The School of Mathematics

THE UNIVERSITY
of EDINBURGH

SI_nR Covid-19 Modeling based on Gamma, Log-normal, and Weibull Generation Interval Distributions

by

Annette Bell

Dissertation Presented for the Degree of
MSc in Operational Research with Data Science

August 2022

Supervised by
Dr John Dagpunar

Abstract

The Susceptible-Infected-Recovered (SIR) compartmental model is used in epidemiology to identify and categorize members of a population based on their status with regards to a disease. Less studied variations of this problem are the SI_nR and SI_nRS models. These models, which have applications in latent infection and varying transmission rates, will be used on three different generation interval—the time between primary exposure and secondary infection—distributions: gamma, log-normal, and Weibull. The distributions are ultimately tested against one another to see not only which provides most realistic data, but how these data-sets interact.

Acknowledgments

Thank you to my academic advisor Dr. Dagpunar for your guidance and support. Thank you to my sister Alexi for being my rock throughout my academic and personal pursuits. Finally, thank you to my parents who have fostered my intellectual curiosity and inspired me to pursue this masters in Edinburgh.

Own Work Declaration

I declare that I composed this thesis myself and that any work that is not my own is explicitly stated. This work has not been submitted for any other degree or qualification.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background

Novel disease SARS-CoV-2, commonly known as COVID-19, was first reported in Wuhan, China in December of 2019. By January 29, 2020, the United Kingdom had its first confirmed case. Two days later, the World Health Organization (WHO) declared a public health emergency of international concern [9]. By March 23, 2020, the United Kingdom implemented its first official lockdown. Within the first year of COVID-19, more than 65 million cases occurred world-wide with 1.5 million resulting in death [39]. Of these cases, 2.5 million occurred in the United Kingdom resulting in 75,000 deaths [41]. By December of 2020, the first COVID-19 vaccine was approved, and individuals began the immunization process [9].

Throughout the pandemic, public health and policy officials turned to epidemiologists in order to better understand the disease trends. Many of these epidemiological predictions aided in the government's decisions to implement non-pharmaceutical interventions—actions individuals can take, excluding medicine and vaccines, to decrease the spread of a disease. Interventions occurred by way of mask wearing, lockdown, quarantine, or isolation among others.

Since its first appearance, scientists fervently examined all aspects of COVID-19 to better understand its potential impact on the community. To understand the potential future trajectory of the disease, some scientists utilized compartmental models. A compartmental model in epidemiology examines a population of interest and places them in one of many labelled compartments identifying their status. One of the most well-known compartmental epidemiological models is the Susceptible-Infected-Recovered (SIR) model. The SIR model places a population into one of three compartments: susceptible, infected, or recovered.

The susceptible compartment represents all the individuals that have not contracted a disease but are at risk or susceptible. People shift to the infected compartment once exposed to the disease. The infected compartment refers to the people currently infected also known as *prevalence*. This is important as the infective compartment does not measure *incidence*—the number of new daily cases. Although appearing to be a subtle difference, this distinction is important to understand who should be in the infected compartment. Finally, individuals move to the final recovered compartment once they are no longer infectious or dead. The recovered compartment is also often referred to as the removed compartment as the most basic SIR model assumes life-long immunity like we typically see in diseases like the Chicken Pox.

Figure 1: SIR Compartmental Model Visual

The SIR model has a multitude of variations like the Susceptible-Exposed-Infected-Recovered (SEIR) and Susceptible-Infected-Susceptible (SIS) models. Based on analysis of epidemiological compartmental models, we selected the SI_nR and SI_nRS models. The I_n in these models account for n infected compartments where each compartment acknowledges its specific transmission rate of an infected person. These models examine the United Kingdom (UK) from the start of COVID-19 in December 2019 until the first vaccination in January 2020.

1.2 Objective

The SI_nR and SI_nRS models add complexity to the base SIR model in order to better capture the real world scenarios. This complexity is seen through the infected compartments. Due to additional infected compartments, the model must now include more transmission rates. In order to better understand the spread of COVID-19 in the UK, we must first estimate the transmission rates β_j for each infected compartment. We must then use these values to simulate an epidemic curve based on the SI_nR and SI_nRS models.

This paper will explain key terminology necessary to understand the extent of the problem as well as highlight previous work and estimations of said parameters. Next, the mathematical equations necessary to calculate the transmission rates will be explained and the algorithmic implementation in python will be highlighted. Then, the results of the transmission rates and compartmental models will be explained and explored.

1.3 Motivation

Modifications to the SIR compartmental model has provided more in-depth and realistic analysis of different diseases throughout history. Although the SI_nR and SI_nRS models already exist, the way in which we analyse them could provide an interesting way to examine data. More specifically, we analyse three different sets of distributions using the same population mean and standard deviation. Then, we apply the data assuming they actually follow one distribution. The purpose of this is to understand if we can use any data and predict similar results. Our findings could help provide some inclination as to whether or not we can actually examine epidemiological models, regardless of their data's distribution, with some degree of confidence through one modelling technique.

2 Previous Literature

The first record of mathematical modelling of infectious disease was in 1760 by Swiss mathematician Daniel Bernoulli [1]. Since then, epidemiologists invented various models to study the course of a disease to include network transmission models [45], individual-based models [31], and compartmental models [19, 37] among others. Compartmental models place individuals of a population into labelled compartments. Some common labels include Susceptible (S), Exposed (E), Recovered or Removed (R), Vaccinated (V), Quarantined (Q), and Dead (D). These models typically use a naming convention based on the flow pattern of the model. For example, if someone wanted to examine the course of infection and vacinations, they might use a Susceptible-Vaccinated-Infected-Recovered (SVIR) model. Since these models can be tailored to model desired attributes of the disease, they were and continue to be popular. The most basic and commonly used of the compartmental models is the SIR model.

2.1 SIR Model

In the early twentieth century, Ronald Ross, William Hamer, and others [5] developed the SIR compartmental model. In the following years, Kermack and McKendrick [22] made importance conclusions that significantly furthered epidemiology. Epidemiologists utilized the SIR model for a myriad of diseases throughout history to include Influenza [19], SARS [32], Ebola [33], and COVID-19 [37] among others. The results provided from the SIR model not only aided the medical community in treating and managing these diseases but also allowed public health and policy officials to advise and mandate non-pharmaceutical interventions if necessary [30]. Many of these epidemiological advancements in the larger medical community were founded on the base SIR model which is a system of three, non-linear, ordinary differential equations.

$$\left.\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= -\beta SI - \mu I \\ \frac{dR}{dt} &= \mu I \end{aligned}\right\} \tag{2.1}$$

where S, I, and R are the number of susceptible, infected, and recovered individuals respectively. μ is the mean ineffective period and β is the transmission rate between infected and susceptible. The SIR model makes five primary simplifying assumptions to model disease.

1. Over the duration of the disease, the total population N is constant and equivalent to $S+I+R$. This means that the SIR model excludes births and deaths from the model (ie. $\frac{dS}{dt}+\frac{dI}{dt}+\frac{dR}{dt}=0$).

2. At the beginning of the outbreak, no one in the population has immunity from the disease. Therefore, everyone that is not infected at the start of the outbreak is in the susceptible compartment.

3. The population is homogeneously mixed—everyone has an equal probability of interacting with one another.

4. The transmission rate β is constant. This means that upon exposure to the disease, an individual is presumed to be as infectious the first day as they are the last day.

5. Once recovered from a disease, a person gains life-long immunity. This signifies that no waning immunity—decrease in antibodies over time—exits.

Although these assumptions allow the SIR model to provide valuable insights for many diseases, COVID-19 has some complexities that the base model is incapable of truly handling. The assumption of a constant transmission rate β is unrealistic. We would expect a person's infectiousness to shift overtime. Upon exposure to the disease, an individual would undergo a period of time where they are carrying the disease but not yet infectious. This is known as a latent period. Once infectious,

their transmission or infectious rate should increase until it hits some peak transmission level and then decrease back to 0. When the individual can no longer transmit the disease, they would then leave the infective compartment and enter the recovered R compartment. Without accounting for the latency period nor change in transmission rate overtime, the SIR model fails to properly account for the true spread of a disease like COVID-19.

Another drawback of this model is that it does not account for waning immunity. Without the presence of waning immunity, the model assumes that once infected, a person cannot be reinfected. This is problematic for continued analysis of COVID-19 as we know individuals can be reinfected. Homogeneous mixing is another limitation of the model as it does not properly relay interactions amongst the population. For example, one would expect school aged individuals to interact significantly more often than a teenager and an elderly person. When we assume that everyone interacts with one another at the same rate, then we may be overestimating interactions between age groups or different business cohorts. Due to these more complex disease proliferation patterns, there are other variations of the SIR model that expand the model to account for latency, waning immunity, birth and death rates, vaccination, quarantine, and more.

2.2 SI_nR Model

The Susceptible-Exposed-Infected-Recovered (SEIR) and SI_nR models provide relief to some of these aspects. The SEIR model introduces the "Exposed" or E compartment to the original SIR model. In this model, susceptible individuals contract the disease and move to the exposed compartment. In

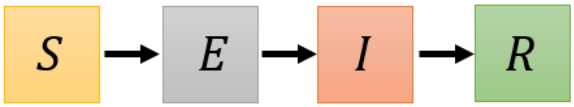

Figure 2: SEIR Compartmental Model Visual

the E compartment, individuals undergo a period of time where they are infected but not yet infectious. They shift to the Infected I and Recovered R compartments upon infectiousness and recovery respectively. The SI_nR model creates n infected compartments to address varying transmission rates throughout the duration of infection.

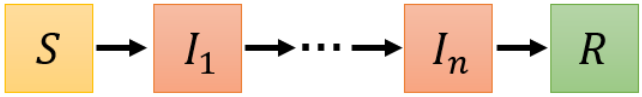

Figure 3: SI_nR Compartmental Model Visual

The system of ordinary differential equations of the SI_nR model are as follow:

$$\left.\begin{aligned} \frac{dS}{dt} &= -\frac{S}{N}\sum_{j=1}^{n}\beta_j I_j \\ \frac{dI_1}{dt} &= \frac{S}{N}\sum_{j=1}^{n}\beta_j I_j - \lambda I_1 \\ \frac{dI_j}{dt} &= \lambda(I_{j-1} - I_j) \quad \forall\ j \in 2, ..., n \\ \frac{dR}{dt} &= \lambda I_n \end{aligned}\right\} \tag{2.2}$$

S is the number of susceptible individuals. I_j is the number of infected people in the jth compartment. There are a total of n infected compartments. I is the total number of infected individuals ($I = \sum_{j=1}^{n} I_j$). R is the number of recovered individuals. N is the total population ($N = S + I + R$). β_j is the transmission rate of the jth compartment. The holding period λ is the period of time an individual

is expected to spend in each infected compartment. The holding rate is the expected infectious time in days and is the inverse of the holding period λ^{-1}.

Like the SEIR model, the SI_nR model distinguishes between infected and infectious. Infected compartments include individuals who have the disease regardless of whether or not they can transmit the disease to someone else. Therefore, all I_j for $j \in 1, ..., n$ compartments are infected compartments. Infected compartments are deemed infectious if their corresponding transmission rates are greater than 0 ($\beta_j > 0$).

The SI_nR model in its two simplest forms ($n = 1$ and $n = 2$) are compared to the SIR and SEIR models respectively. The SI_1R model is equivalent to the the SIR model. When $n = 2$, the SI_2R model is typically compared to the SEIR model, where the first infective compartment I_1 is compared to the exposed compartment and the second I_2 is viewed as the infected compartment. In this case, we would expect the transmission rate of the first compartment to equal 0 ($\beta_1 = 0$) as the individual is exposed but not yet infectious. The second compartment's transmission rate would be some value greater than 0 ($\beta_2 > 0$), expressing the transmission rate over the course of the entire infectious period. In this case, the SI_2R model assumes a constant transmission like the base model. Once the compartment size exceeds two ($n > 2$), the SI_nR model is a generalized SEIR model where each compartment j has its respective transmission rate β_j that can be any value greater than or equal to 0. The SI_nR model addresses the issue of latency and transmission rate that the basic SIR model does not properly account for with regard to Covid. The SI_nR model continues to assume a constant population, that no one has initial immunity, a homogeneously mixed community, and life-long immunity.

The SI_nR model, used in studies related to influenza [19] and measles [23] among other diseases, will be used in this study to model COVID-19 in the United Kingdom. Since this model introduces greater complexity with β_j, the SI_nR model requires additional mathematical calculations for the now varying transmission rate. For this reason, there are a number of parameters that must be understood to properly model the disease.

2.3 Definitions of Key Epidemiological Parameters

2.3.1 Basic Reproduction Number (R_0)

The basic reproduction number, also referred to as the basic reproduction rate or ratio, was first conceptualized for demography in 1925 to count the number of newborns [20]. This concept was adapted in 1980 to count infectious cases for epidemiology [20]. The basic reproduction number (R_0) is the number of secondary cases produced by an average infected person in a fully susceptible population [36]. The basic reproduction number assumes a well-mixed community with equal opportunity for interaction [36]. Additionally, the basic reproduction number assumes that no non-pharmaceutical interventions, like mask wearing or social distancing, exist.

As the disease progresses and more individuals depart the susceptible population, the basic reproduction number becomes a less suitable parameter to model the disease. For this reason, studies may use the effective reproduction number, also referred to as the time dependent reproduction number, R_t. In this instance, the effective reproduction number R_t is the transmission potential given a specific time period t in the pandemic [2]. As such, R_t and R_0 differ in their uses. For example, to measure a population three months after primary infection, then R_t where $t = 3$ would be preferred as the population would not be wholly susceptible. As we are measuring from the beginning of the disease, R_0 will be used.

The basic reproduction number is important to epidemiology as it has many implications for how a communicable disease will progress. If the basic reproduction number exceeds one ($R_0 > 1$), then we expect the disease to spread [11]. However, if the basic reproduction number is less than one ($R_0 < 1$), then we expect it to eventually end [11]. If R_0 is equivalent to 1, then the disease is endemic—the population replenishes the number of susceptible individuals. For this reason, the lower the basic reproduction number the better.

The basic reproduction number within the SIR model is measured as the likelihood that an individual will infect a susceptible individual at a transmission rate β multiplied by the expected duration of infection $\frac{1}{\mu}$:

$$R_0 = \frac{\beta}{\mu}, \tag{2.3}$$

Given a more complex interaction due to the increased number of compartments, the basic reproduction number R_0 in the equation above is modified to account for multiple infectious compartments. The basic reproduction number will therefore be equivalent to the sum of transmission rates of each infectious compartment multiplied by the mean time in each infected compartment, or holding rate.

$$R_0 = \frac{\sum_{j=1}^{n} \beta_j}{\lambda}, \tag{2.4}$$

Although the holding period will be solved for in the non-linear optimization problem in Section 3.2, λ is initially assumed to be the number of compartments n multiplied by the expected duration of the infection $\frac{1}{\mu}$. In addition to the basic reproduction number, the generation interval provides additional insights into the growth dynamics of COVID-19 and other diseases.

2.3.2 Generation Interval (τ)

There are two phases over the course of most infections. After an individual is exposed to and contracts the disease, they undergo a latent period in which the individual is infected with the disease but is not yet infectious. After the latent period ends, an individual becomes infectious. The time between primary exposure and secondary infection is known as the generation interval, or generation time (τ) [44]. The figure below is from a 2020 study [44] that estimated the generation interval of COVID-19 from contact tracing data and illustrates the course of infection for a infector infectee pair.

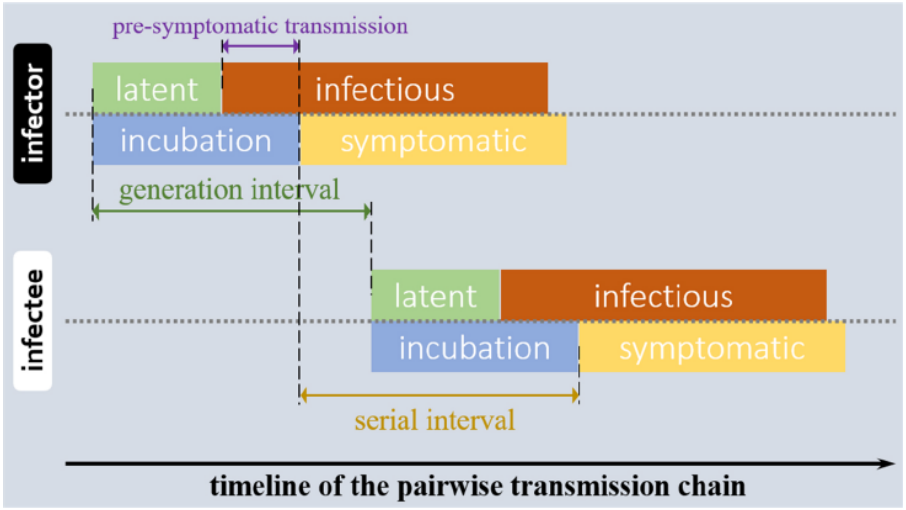

Figure 4: Course of Disease from Infector to Infectee [44]

The generation interval measures prevalence—number of people presently infected. The best way to measure prevalence is to random sample a population with periodic (whether daily, weekly, or monthly) COVID-19 testing. The current gold standard test is the Polymerase Chain Reaction (PCR) test; however, some studies use the quicker and less expensive lateral flow antigen test. Through random sampling, the data captures a more realistic view of how many people are currently infected. This random sampling is critical as asymptomatic individuals are unlikely to test given the fact they feel fine. This means that there is a better chance at capturing both the symptomatic who might test when they suspect it is COVID-19 as well as asymptomatic individuals unlikely to test in such cases.

The generation interval is more difficult to measure as initial exposure is typically unobserved. As such, some studies [24] utilize the serial interval as a proxy for generation interval data. The serial interval is the time between the infector's symptomatic phase and the infectee's symptomatic phase [18, 24, 44]. Although the generation interval and serial interval measure different periods of an infection, they typically have similar mean and standard deviations, providing a reasonable

substitution for either of these parameters.

There have been a variety of studies conducted for both the serial interval and generation intervals [16, 18, 24, 44] for Covid. Some studies used real data while others used simulated data. From these studies of the generation interval for COVID-19, many data studies [16, 24, 27] have determined that Gamma, Weibull, and or Log-normal distributions are most representative of the data. Each of these distributions, will be examined in this study.

3 Model Design and Implementation

Before implementing the compartmental model, we first selected values for the basic reproduction number R_0 and the generation interval τ from preliminary analysis. We then estimated the transmission rates β_j and holding period λ for each of the n compartments using a non-linear optimization problem based on the expected infectious curve $\beta(\tau)$. All calculations and modelling were implemented through python algorithms.

3.1 Selection of Key Epidemiological Parameters

3.1.1 Basic Reproduction Number

Previous literature [2, 26] states that the basic reproduction number for diseases, including COVID-19, varies depending on location. In the figure below, a 2020 study [26] on the basic reproduction number of COVID-19 in European countries illustrated its findings by country. The data was drawn from the 27 countries of the European Union. The study extracted information related to differences in report dates, air traffic, mobility via cell phone data, and more with the baseline date of January 13, 2020. Utilizing machine learning practices, the study then smoothed the day fluctuations through a seven-day moving average. Figure 5 highlights the study's 2020 findings on European countries varying mean

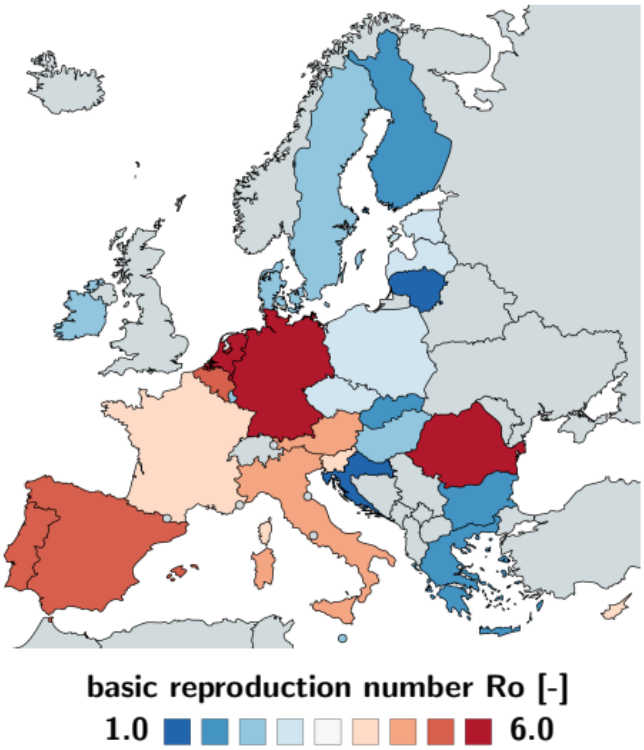

Figure 5: Estimated Basic Reproduction Numbers of European Countries from 2020 Study [26]

basic reproduction values. The lowest mean basic reproduction rate the study predicted originated from Lithuania with a value of 0.91 and the highest from Germany with 6.33. The next table below displays various studies' estimates for the basic reproduction number R_0 of the United Kingdom. We see that the lowest predicted basic reproduction number is 2 while the highest value is 4.8.

Source	R_0 Estimate
Jit M et al [21]	2 (1.9-2.1)
Lonergan M, Chalmers J [28]	2.1 (1.8-2.3s)
Lourenço J, Paton R, Ghafari [29]	2.25 or 2.75
Brett T, Rohani P [7]	2.3
Ferguson N, Laydon D, Nedjati-Gilani G, et al [13]	2.4 (2.0-2.6)
Lonergan M, Chalmers J [28]	2.6 (2.4-2.9)
Althouse B, Wenger E, Miller J [4]	2.6
European Center for Disease Prevention and Control [15]	3.28
Chen X, Dong Y, Xiaoyue Y [42]	4.8 (4.7-4.9)

Table 1: Estimated Basic Reproduction Number for the United Kingdom used from Other Studies.

The majority of these predictions; however, fell between 2 and 2.6 for R_0. This model will implement a basic reproduction value of 2.3 which consistent with many other studies [7, 13, 25, 29, 43]. A 2020 study [7] related to COVID-19 in the United Kingdom and containment measures possible by the governing body also used this value in their study. Due to various studies supporting this estimate, we deemed it to be a sufficient estimate.

3.1.2 Generation Interval

As previously mentioned in Section 2.3, numerous studies were conducted to calculate the generation interval of different locations. Most of these studies determined a gamma, Weibull, or log-normal distribution to be the best fit of the generation interval for Covid. Some of these studies' findings are shown in the table below.

Source	Location	Distribution	Mean	StD	Time Period
Challen et al., 2021 [10]	UK	Gamma	4.9	2	Jan - Mar2020
Li et al., 2020 [24]	China	Gamma	4.81	2.52	21Jan - 29Feb2020
Li et al., 2020 [24]	China	Lognormal	4.72	2.64	21Jan - 29Feb2020
Park et al., 2020 [34]	China	Lognormal	1.54	0.37	21Jan - 8Feb2020
Ferretti et al., 2020 [14]	Various	Weibull	5	1.9	Dec2019 - Feb2020
Li et al., 2020 [24]	China	Weibull	4.82	2.38	21Jan - 29Feb2020

Table 2: Generation Intervals Calculated from Other Studies.

Given the generation interval statistics, this project simulated data for each of the three distributions. To best compare the results, each distribution followed Challen's 2020 UK study [10] with mean of 4.9 and standard deviation of 2. This distribution assumed independence between transmission and symptom onset [10] and was estimated based on information from January to March 2020. Furthermore, this distribution was based on re-sampled serial intervals as a predictor.

To create the data for each distribution, we calculated the specific parameters using the population mean $\mu = 4.9$ and standard deviation $\sigma = 2$. The gamma distribution relied on the shape α and inverse scale θ parameters which were calculated from the following equations.

$$\alpha = \frac{\mu^2}{\sigma^2} \qquad\qquad \theta = \frac{\sigma^2}{\mu^2}$$

The generation interval was then simulated in python from Gamma($\alpha = 6.003, \theta = 0.167$). Next, we created the data of a log-normal distribution. The log-normal has the mathematical properties that $\mathcal{X} \sim \text{Lognormal}(\mu = 4.9, \sigma^2 = 2^2)$ and $\ln(\mathcal{X}) \sim \mathcal{N}(\mu_n, \sigma_n^2)$. The relationship between the mean and variance for the log-normal and normal distributions are below.

$$\mu = e^{\left(\mu_n + \frac{\sigma_n^2}{2}\right)} \qquad\qquad \sigma^2 = [e^{\sigma_n^2} - 1]e^{(2\mu_n + \sigma_n^2)}$$

Through manipulating these equations, we were able to solve for the mean and variance of the normal distribution.

$$\sigma_n^2 = \ln\left(\frac{\sigma^2}{\mu^2} + 1\right) \qquad \mu_n = \ln(\mu) - \frac{\sigma_n^2}{2} \tag{3.1}$$

Using the population mean and standard deviation from Challen's study, we found $\sigma_n^2 = 0.154$ and $\mu_n = 1.512$. Therefore, the shape was $\sigma_n = 0.393$ and the scale was $e^{\mu_n} = 4.537$. The shape and scale were input into a SciPy log-normal function and the data was generated. For the final distribution, we found the interaction between the mean and standard deviation with the shape k and scale γ parameters

$$\mu = \gamma\Gamma\left(1 + \frac{1}{k}\right) \qquad \sigma^2 = \gamma^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right]$$

We then modified the Weibull equations above to calculate the shape k parameter and then solve for the scale γ.

$$\frac{\Gamma\left(1 + \frac{2}{k}\right)}{\Gamma^2\left(1 + \frac{1}{k}\right)} = \frac{\sigma^2}{\mu^2} + 1 \qquad \gamma = \frac{\mu}{\Gamma\left(1 + \frac{1}{k}\right)}$$

We used a solver for k in python and found that it was equal to 2.635. The scale was therefore 5.514. As each of these variables were simulated and therefore an approximation of their respective predicted distributions, their values were not exactly the same. As shown in Figure 6, the gamma generation interval had a sample mean of 4.976 and standard deviation of 1.991. The log-normal and Weibull generation interval data sets simulated had sample means of 4.906 and 4.903 as well as standard

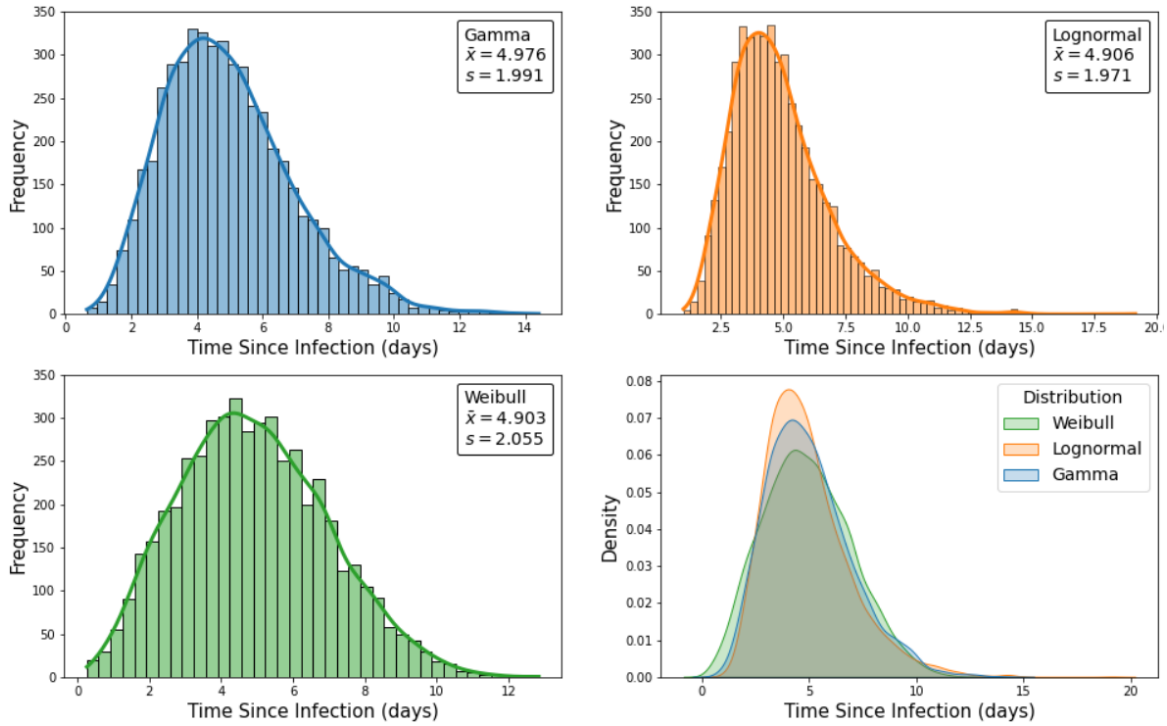

Figure 6: Generation Interval Simulated from Gamma (blue), Log-normal (orange), and Weibull (green) Distributions and their Estimated Kernel Densities

deviations of 1.971 and 2.055 respectively. Each of these distributions were simulated in python using a seed of 1234 with 5000 observations. We selected random sampling as opposed to the pseudo-random hypercube as both were suitable options and the random sampling produced sample means and standard deviations that were close to the original population mean of 4.9 and standard deviation of 2.

The graph in the bottom right section depicts the kernel density estimations for each of the distributions to better visualize how similar these data sets were. All three of the distributions covered large overlapping areas. From this graph, we notice that the Weibull had a slightly larger span of generation times while the log-normal had the most dense generation interval near 4.9. The graph in the bottom right more explicitly illustrates that, although simulated from the same population mean, standard deviation, seed, and number of observations, these distributions did vary.

3.2 Expected Infectious Curve, $\beta(\tau)$

3.2.1 Non-Linear Optimization

A Poisson distribution is a discrete probability distribution that measures the number of occurrences of an event in a specified period given a mean rate λ of the event happening. The distribution was used to estimate disease occurrence and measured the probability of observing $j-1$ events in time period 0 to τ given rate $\lambda > 0$.

$$p(\tau) = \frac{e^{-\lambda\tau}(\lambda\tau)^{(j-1)}}{(j-1)!}, \tag{3.2}$$

$$\lambda > 0, \tag{3.3}$$

Each of the compartments was given by j where $j \in 1, .., n$. The holding rate was λ and was a constant. The Poisson distribution was then introduced to the expected infectious curve through its relationship with each infectious rate. The transmission rate β_j with probability that there were $j-1$ events in a Poisson process of rate λ in the time interval from 0 to τ was therefore:

$$\beta(\tau) = \sum_{j=1}^{n} \beta_j \frac{e^{-\lambda\tau}(\lambda\tau)^{(j-1)}}{(j-1)!}, \tag{3.4}$$

The expected infectious curve was used for the generation interval data τ regardless of the distribution it followed. This formulation was important as it not only determined the expected infectiousness of a person who had been infected from time 0 to τ, but also was the base function to determine each of the β_j values. Although this function was important, it lacked clear relationships between transmission rates to quickly solve the overall function. For this reason, the transmission rates β_j for each compartment j were interpreted in terms of their weighted sums w_j. The transmission weight function $W(\tau)$ was therefore the sum of the weights.

$$w_j = \frac{\beta_j}{\sum_{j=1}^{n} \beta_j}, \tag{3.5}$$

However, to calculate the weights, the Poisson distribution (Eqn. 3.2) had to be changed from the probability mass function to a probability density function (PDF). To change the discrete to a continuous function, we multiplied $p(\tau)$ by rate λ. The transmission weight function was therefore the sum of the weights times the probability distribution function.

$$W(\tau) = \sum_{j=1}^{n} \frac{\beta_j}{\sum_{j=1}^{n} \beta_j} \frac{e^{-\lambda\tau}(\lambda\tau)^{(j-1)}\lambda}{(j-1)!}$$

$$W(\tau) = \sum_{j=1}^{n} w_j \frac{e^{-\lambda\tau}(\lambda\tau)^{(j-1)}\lambda}{(j-1)!}, \tag{3.6}$$

By introducing weights of the infectious rates, we gained mathematical properties previously unknown. The weights must sum to one and each weight is bounded by 0 and 1.

$$\sum_{j=1}^{n} w_j = 1, \tag{3.7}$$

$$0 \leq w \leq 1 \qquad \forall\ j \in 1, .., n, \tag{3.8}$$

The relationship between the sum of the weights was further confirmed by the properties of the PDF. The PDF used had properties that it is non-negative for all values of τ and the probability over the entire space is equivalent to 1.

$$\int_0^\infty \frac{e^{-\lambda\tau}(\lambda\tau)^{(j-1)}\lambda}{(j-1)!}\, d\tau = 1$$

$$\int_0^\infty \sum_{j=1}^{n} w_j \frac{e^{-\lambda\tau}(\lambda\tau)^{(j-1)}\lambda}{(j-1)!}\, d\tau = \sum_{j=1}^{n} w_j = 1$$

To estimate the transmission weights w_j , we first calculated the maximum likelihood $\mathcal{L}$ as a function of generation intervals τ and weights w_j.

$$\mathcal{L}(\tau_1, ..., \tau_m; w_1, ..., w_n) = \prod_{i=1}^{m} \sum_{j=1}^{n} \left(w_j \frac{e^{-\lambda\tau_i}(\lambda\tau_i)^{(j-1)}\lambda}{(j-1)!} \right),$$

Then, we found the log-likelihood by taking the log of the maximum likelihood.

$$\log\left(\mathcal{L}(\tau_1, ..., \tau_m; w_1, ..., w_n)\right) = \sum_{i=1}^{m} \left(\log\left(w_1 + w_2 \frac{(\lambda\tau_i)^1}{1!} + ... + \frac{(\lambda\tau_i)^{n-1}}{(n-1)!} \right) - \lambda\tau_i + \log\lambda \right), \tag{3.9}$$

With the log-likelihood calculated, we solved for the weights w_j and transition rate between infectious compartments (λ). The non-linear optimization problem maximized the log-likelihood function given the weight constraints (Eqns. 3.7 and 3.8) and the lambda constraint (Eqn. 3.3).

$$\begin{aligned} \max \quad & \sum_{i=1}^{m} \left(\log\left(w_1 + w_2 \frac{(\lambda\tau_i)^1}{1!} + ... + \frac{(\lambda\tau_i)^{n-1}}{(n-1)!} \right) - \lambda\tau_i + \log\lambda \right) \\ \text{s.t.} \quad & \sum_{j=1}^{n} w_j = 1 \\ & 0 \leq w \leq 1 \qquad \forall\ j \in 1, .., n \\ & \lambda > 0 \end{aligned} \tag{3.10}$$

The nonlinear optimization problem above solved for the weights w_j for all $j \in 1, ..., n$ and the transition rate λ. With these values, we used the relationships between R_0, β_j, and w_j found in Equations 2.4 and 3.5 to calculate each of transmission rates.

$$R_0 = \frac{\sum_{j=1}^{n} \beta_j}{\lambda} \quad \therefore \quad \sum_{j=1}^{n} \beta_j = R_0\lambda$$

$$w_j = \frac{\beta_j}{\sum_{j=1}^{n} \beta_j} \quad \therefore \quad \sum_{j=1}^{n} \beta_j = \frac{\beta_j}{w_j},$$

$$\beta_j = R_0\lambda w_j \tag{3.11}$$

Given the mathematical calculations above for the transmission rates and holding period, we were then able to apply the estimated parameters to the compartmental models.

3.3 Epidemic Compartmental Models

3.3.1 SI_nR Model Implementation

The SI_nR model, as displayed in Equation 2.2, was slightly amended to account for contact.

$$\left.\begin{aligned} \frac{dS}{dt} &= -\frac{S}{N}\sum_{j=1}^{n}\beta_j I_j c \\ \frac{dI_1}{dt} &= \frac{S}{N}\sum_{j=1}^{n}\beta_j I_j c - \lambda I_1 \\ \frac{dI_j}{dt} &= \lambda(I_{j-1} - I_j) \quad \forall\ j \in 2, ..., n \\ \frac{dR}{dt} &= \lambda I_n \end{aligned}\right\} \tag{3.12}$$

In this modified system of equations, the contact rate c represents the likelihood of interaction between individuals in the population. By introducing contact rate c into the equation, $c * \beta_j$ is better defined as the probability of contact multiplied by the conditional probability given contact. When $c = 1$, then everyone is considered to interact normally and implies that no non-pharmaceutical interventions were imposed on the population and can be viewed as the "worst case scenario". $c = 0$ implies that every single person is completely isolated and therefore can not transmit the disease to anyone else.

One study [38], which implemented a Susceptible-Quarantined-Infected-Removed (SQIR) COVID-19 model, estimated transmission and contact rates based on 32,583 laboratory confirmed COVID-19 Wuhan cases from December 8, 2019 until March 8, 2020. The study applied this model to Italy, the United Kingdom, and the United States. With regards to the United Kingdom, they found that the initial transmission rate β, when no non-pharmaceutical intervention was present ($c = 1$), was 4.6. The study estimated that the initial nationwide lockdown on March 23, 2020 decreased the contact rate to $c = 0.5$. Furthermore, the study also found that the lockdown strengthened its mandates on March 31, 2020, setting the contact rate to $c = \frac{1}{3}$. The table below highlights estimated contact rates based on several non-pharmaceutical techniques.

Source	Non-Pharmaceutical Intervention	c Estimate
Wang et all. [38]	Lose UK Lockdown	0.5
Wang et all. [38]	Strict UK Lockdown	$0.\bar{3}\bar{3}$
Eikenberry et all [12]	Mask	0.85

Table 3: Estimate of Contact Rate c by Non-Pharmaceutical Intervention

Given these varying contact rates of non-pharmaceutical interventions, the c value in the SI_nR model could be changed to examine how these safety measures affect the transmission of COVID-19. With the introduction of c, the model now assumes a constant contact rate of some value c where $0 \geq c \geq 1$. One of the biggest pitfalls of the SIR and SI_nR models is that it excludes waning immunity present in Covid. As such, we also modelled the SI_nRS model.

3.3.2 SI_nRS Model Implementation

The primary difference between the two models is that the SI_nRS model accounts for waning immunity. As such, individuals who previously recovered from COVID-19 will return to the susceptible population at the waning immunity rate h.

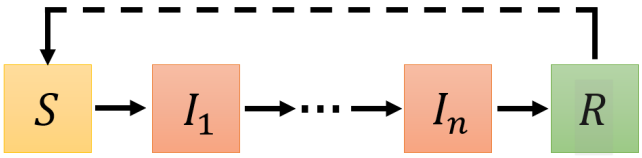

Figure 7: SI_nRS Compartmental Model Visual

The SI_nRS compartment model system of equations is therefore:

$$\left.\begin{aligned} \frac{dS}{dt} &= -\frac{S}{N}\sum_{j=1}^{n}\beta_j I_j c + hR \\ \frac{dI_1}{dt} &= \frac{S}{N}\sum_{j=1}^{n}\beta_j I_j c - \lambda I_1 \\ \frac{dI_j}{dt} &= \lambda(I_{j-1} - I_j) \quad \forall\ j \in 2, ..., n \\ \frac{dR}{dt} &= \lambda I_n - hR \end{aligned}\right\} \tag{3.13}$$

where h is the waning immunity rate with the bound $h > 0$. R is the number of disease resistant individuals. This compartment is called the resistant compartment to acknowledge that people can become reinfected as their immunity wanes. Like the SI_nR model, this model assumes a constant population, no initial immunity, and a homogeneously mixed population. To estimate and model the transmission rates and compartmental models, we created and used python algorithms.

3.4 Model Strategy

3.4.1 Calculating the Transmission Rates

Three python functions were created to solve for the unknown of the weights w_j, holding period λ, and transmission rates β_j. These functions included the objective, constraint, and the solver functions. The solver algorithm used the minimize function within the SciPy library to solve the non-linear optimization problem. As such, the code was solved using the minimized objective.

$$\min \quad -\left(\sum_{i=1}^{m}\left(\log\left(w_1 + w_2\frac{(\lambda\tau_i)^1}{1!} + ... + \frac{(\lambda\tau_i)^{n-1}}{(n-1)!}\right) - \lambda\tau_i + \log\lambda\right)\right) \tag{3.14}$$

The objective function modeled from Equation 3.14, required initial inputs of the weights, holding period, and generation interval observations. It then returned the objective value given said inputs.

Algorithm 1: Function that calculates the objective value given the two inputs: a list of the weights and holding period as well as the generation interval times.

Function: objective $([w_1, .., w_n, \lambda], [\tau_1, ..., \tau_m])$;
Input: (1) List of weights and holding time: $[w_1, .., w_n, \lambda]$
(2) List of generation interval times: $[\tau_1, ..., \tau_m]$
Output: Calculated objective value: *obj_val*

obj_val = 0
for $i = 1$ *to* m **do**
  $u_{calc} := 0$
  for $j = 1$ *to* n **do**
    $u_{calc} = u_{calc}\ w_j \frac{(\lambda\tau_i)^{j-1}}{(j-1)!}$
  end
  $obj\_val = obj\_val + \log(u_{calc}) - \lambda\tau_i + \log(\lambda)$
end
return *obj_val*

The constraint function based on Equation 3.7, took the weights and holding period as inputs. It then calculated the sum of the weights. Due to the way SciPy's library functions, the algorithm had to be written in a way it would understand. For this reason, the algorithm returns 1-$\sum_{j=1}^{n} w_j$. SciPy then clarifies that the value returned must equal 0. As such, this algorithm sets the sum of the weights equal to one.

Algorithm 2: Function that specifies the constraint that the sum of the weights must equal one.

Function: constraint1 $([w_1, .., w_n, \lambda])$;
Input: List of the weights and holding period $[w_1, .., w_n, \lambda]$
Output: 1-$\sum_{j=1}^{n} w_j$

constraint = 1
for $j = 1$ *to* n **do**
  constraint= *constraint* $- w_j$
end
return *constraint*

The objective and constraint functions were implemented in the final of the three functions which was used to calculate the weights, holding time, and transmission rates. The *solver* function minimized the negative log-likelihood (Eqn. 3.14). The function had four main components.

1. The first section (lines 2-14) set the initial estimates for the weights and holding period $[w_1, ..., w_n, \lambda]$. As this problem was a multi-dimensional optimization problem, it was important to supply the algorithm with good initial estimates to decrease the solve time. Though experimentation, we found the proper initial estimates for the gamma, Weibull, and log-normal distributions. The formal analysis can be found in Section 4.1.1 subsequently.

Distribution	λ **Estimate**	**Transmission Weight Estimates** w_j
Gamma and Log-normal	$\left(\frac{\bar{x}}{s}\right)^2$	$w_j = 0 \; \forall \; \{j \in 1, ..., n\} \notin \{\text{rounddown}(\alpha), \text{roundup}(\alpha)\}$ $w_{(\text{rounddown}(\alpha))}$=$\alpha - \text{rounddown}(\alpha)$ $w_{(\text{roundup}(\alpha))}$=$1 - w_{(\text{rounddown}(\alpha))}$
Weibull	s	$w_j = \frac{1}{n} \; \forall \; j \in 1, ..., n$

Table 4: Initial Estimates of the Holding Period and the Weights

2. The second section imposed the bounds for the weights and holding period. Regardless of the distribution of the generation interval, the holding period and weights had to follow the boundaries that $\lambda > 0$ and $0 \geq w_j \geq 1 \; \forall \; j \in 1, ..., n$ (Eqn. 3.10). The bound constraints were defined in the variable *bnds*.

3. This section of code (lines 22-23) used Algorithm 2 which stated that the weights must sum to one (Eqn. 3.7). The constraint was assigned to variable *cons*.

4. The fourth section implemented SciPy's minimize function and solved for the weights, holding period, and transmission rates. The minimize function took six inputs. The first input was the objective function created with Algorithm 1. The second input took the initial estimates for the weights and holding period in list format. The third input was the additional arguments that needed to be passed to the objective function. The objective function, in addition to the initial guesses $[w_1, ..., w_n, \lambda]$, needed the generation interval times $[\tau_1, .., \tau_m]$. As such, the additional argument passed to the minimize function was the list of generation interval times. The method in which to minimize the function was the fourth input. Since the equation had constraints and boundaries, we used Sequential Least Squares Programming (SLSQP) as the solver. SciPy's SLSQP method implemented the subroutine by Dieter Kraft. The fifth and sixth inputs took the bounds (*bnds*) and constraints (*cons*) respectively. With the SciPy minimize function completed, the *solution* output the optimized weights and holding time. Equation 3.11 was then used to calculate the transmission rates given the interaction between the basic reproduction number, the optimized holding time, and respective weights. Finally, the function returned the optimized weights w_j^*, holding time λ^*, and transmission rates β_j^*.

Algorithm 3: Function that minimizes the objective subject to the constraint and bounds.

1 **Function:** solver $([\tau_1, ..., \tau_m], R_0, n)$;
Input: (1)List of the generation interval times $[\tau_1, ..., \tau_m]$
(2) Basic reproduction number (R_0)
(3) Number of compartments n
(4) String of the distribution type: *dist_type* $\in$ ["gamma", "lognorm", "weibull"]
Output: (1) List of optimal weights: $[w_1^*, \ldots, w_n^*]$,
(2) Optimal holding period: λ^*
(3) List of optimal transmission rates: $[\beta_1^*, \ldots, \beta_n^*]$

2 **if** *dist_type == "gamma"* ***or*** *dist_type == "lognorm"* **then**
3 $\quad \alpha = \frac{mean([\tau_1, ..., \tau_m])^2}{var([\tau_1, ..., \tau_m])}$
4 $\quad w_{(\text{rounddown}(\alpha))}$=$\alpha - \text{rounddown}(\alpha)$
5 $\quad w_{(\text{rounddown}(\alpha)+1)}$=$1 - (\alpha - \text{rounddown}(\alpha))$
6 $\quad \lambda$= $\frac{mean([\tau_1, ..., \tau_m])}{variance([\tau_1, ..., \tau_m])}$
7 **end**

8 (*Algorithm 3 continues on next page*)

if *dist_type* == *"weibull"* **then**
  for $j = 1$ *to* n **do**
    $w_j = \frac{1}{n}$
  end
  $\lambda = \text{std}([\tau_1, ..., \tau_m])$
end

$b = (0,1)$
$bnds = ()$
for $j = 1$ *to* n **do**
  $bnds = bnds + (b,)$
end
$b_\lambda = (0.000000001,\text{None})$
bnds =bnds+(b_λ,)

con1 = {'type': 'eq','fun': **constraint**}
$cons = ([\text{con1}])$

solution= minimize(*objective*, $[w_1, \ldots, w_n, \lambda]$, args=($[\tau_1, ..., \tau_m]$), method= '$SLSQP$', bounds= $bnds$, constraints= $cons$)
$[w_1^*, \ldots, w_n^*, \lambda^*]$= solution.x
for $j = 1$ *to* n **do**
  $\beta_j^* = w_j \lambda R_0$
end

return $[w_1^*, \ldots, w_n^*], \lambda^*, [\beta_1^*, \ldots, \beta_n^*]$

With the optimized weights, holding period, and transmission rates, the epidemic model was constructed.

3.4.2 Calculating the Course of Covid

The $\text{SI}_\text{n}\text{R(S)}$ algorithm accepted seven inputs and returned the rate of change for each compartment at some specified time. These inputs included the initial number of people in each compartment y_{t0}, the time period we wanted to model over in days $[0, .., t]$, the total population N, contact rate c, and waning immunity rate h. The final two inputs were the optimal transmission rates $[\beta_1^*, ..., \beta_n^*]$ and holding period λ^* calculated earlier using Algorithms 1, 2, and 3. This function worked for both the $\text{SI}_\text{n}\text{R}$ and $\text{SI}_\text{n}\text{RS}$ models. When the recovery period $h = 0$, then the function operated as the $\text{SI}_\text{n}\text{R}$ model as no one was capable of recovery. However, if $h > 0$, then waning immunity existed and individuals eventually returned to the susceptible population.

Algorithm 4: Function returns how many people are in each compartment on a specified day of the SI_nR or SI_nRS compartmental model.

Function: $\underline{SI_nR}$ $(y_{t0},\ [0,...,t],\ N,\ c,\ h,\ [\beta_1^*,...,\beta_n^*],\ \lambda^*)$;
Input: (1) List of people in each compartment on day 0: $y_{t0} = [S(0), I_1(0), ..., I_n(0), R(0)]$,
(2) List of days from 0 to t: $[0,...,t]$
(3) Total population N
(4) Contact rate c
(5) Waning immunity rate h
(6) List of optimal transmission rates of each compartment j: $[\beta_1^*,...,\beta_n^*]$
(7) Optimal holding period λ^*
Output: A list of the rates of change for each compartment: *pop_status*

$S = S(0)$
$I = I_1(0) + ... + I_n(0)$
$R = R(0)$

$infected := 0$
for $j = 1$ *to* n **do**
  $infected = infected + \beta_j * I_j$
end

$\frac{dS}{dt} = -\frac{S}{N} * infected * c + hR$
$\frac{dI_1}{dt} = \frac{S}{N} * infected * c - \lambda^* I_1$
for $j = 2$ *to* n **do**
  $\frac{dI_j}{dt} = \lambda^*(I_{j-1} - I_j)$
end
$\frac{dR}{dt} = \lambda^* I_n - hR$
$pop\_status = [\frac{dS}{dt}, \frac{dI_1}{dt}, ..., \frac{dI_n}{dt}, \frac{dR}{dt}]$
return *pop_status*

The second function SI_nR_Calc utilizes the odeint function within the SciPy library which integrates a system of ordinary differential equations. With regards to the SI_nR and SI_nRS models, this function calculates the percentage of the population in each compartment from time 0 to t. This algorithm requires the same seven inputs as Algorithm 4. Algorithm 5 uses the odeint function which requires four inputs to run.

1. The first input is the function, created and returned with Algorithm 4, to be examined.
2. The second input is the initial conditions for each of the compartments. This includes the initial people in the susceptible, infected, and recovered compartments.
3. The third argument is the sequence of time points we wish to solve for. For example, if you only want to examine the first 30 days, the the time would include points from 0 to 29.
4. The fourth and final input is the additional arguments to pass to the function of Algorithm 4s. For this reason, the fourth input will be a list of the total population N, contact rate c, waning immunity rate h, optimal transmission rates β_j^*, and optimal holding period λ^*.

The algorithm returns the percentage of the population in each category over the designated time period.

Algorithm 5: Function returns the percentage of the population in each compartment each day from day 0 to t of the SI_nR or SI_nRS compartmental model.

Function: $\underline{SI_nR\_Calc}$ $(y_{t0},\ [0,...,t],\ N,\ h,\ [\beta_1^*,...,\beta_n^*],\ \lambda^*)$;

Input: (1) List of people in each compartment on day 0: $y_{t0} = [S(0), I_1(0), ..., I_n(0), R(0)]$,
(2) List of days from 0 to t: $[0,...,t]$
(3) Total Population N
(4) Contact Rate c
(5) Waning Immunity period h
(6) List of optimal transmission rates of each compartment j: $[\beta_1^*,...,\beta_n^*]$
(7) Optimal holding period λ^*

Output: A list of lists that have the percentage of the population for each compartment over time 0 to t: $\left[\left[\frac{S(0)}{N},...,\frac{S(t)}{N}\right],\left[\frac{I(0)}{N},...,\frac{I(t)}{N}\right],\left[\frac{R(0)}{N},...,\frac{R(t)}{N}\right]\right]$

ret = odeint($SInR$, y_{t0}, $[0,...,t]$, args=(N, c, h, $[\beta_1^*,...,\beta_n^*]$, λ^*))

$S = ret$.T[0]

I = sum(ret.T[1:-1])

$R = ret$.T[-1]

return $\left[\left[\frac{S(0)}{N},...,\frac{S(t)}{N}\right],\left[\frac{I(0)}{N},...,\frac{I(t)}{N}\right],\left[\frac{R(0)}{N},...,\frac{R(t)}{N}\right]\right]$

With the python functions created, we ran the models and found our results.

4 Results

4.1 Understanding Compartment Size n Relationships

To select the compartment size to use for the rest of our results, we first had to analyse the distributions. As highlighted in Section 3.4, the process to solve the transmission rates was multidimensional and computationally expensive. To reduce overall run time for each of the three distributions, we first found the preferred initial estimates for the transmission weights and holding period. We were then able to better examine a range of values and ultimately select a compartment size that was sufficiently large.

4.1.1 Initial Estimate Selection

For each of the three distributions, we began with initial estimates of $w_j = \frac{1}{n} \; \forall \; j \in 1, .., n$ and $\lambda = n\sigma$ where σ was the mean infective period. One study [8] found that the mean infective period varied significantly for COVID-19. The study found that asymptomatic cases had mean infective period of 6.5–9.5 days while symptomatic cases lasted an average of 13.4 days from symptom onset [8]. Based on these findings, we set the estimated infective period to $\frac{1}{14}$ [8].

From these initial estimates, we found that the expected transmission curve of the gamma distributed data became repetitive after the compartment size n increased to six. Based on these findings, we then analysed similarities between compartment sizes six through 25, primarily focusing on the

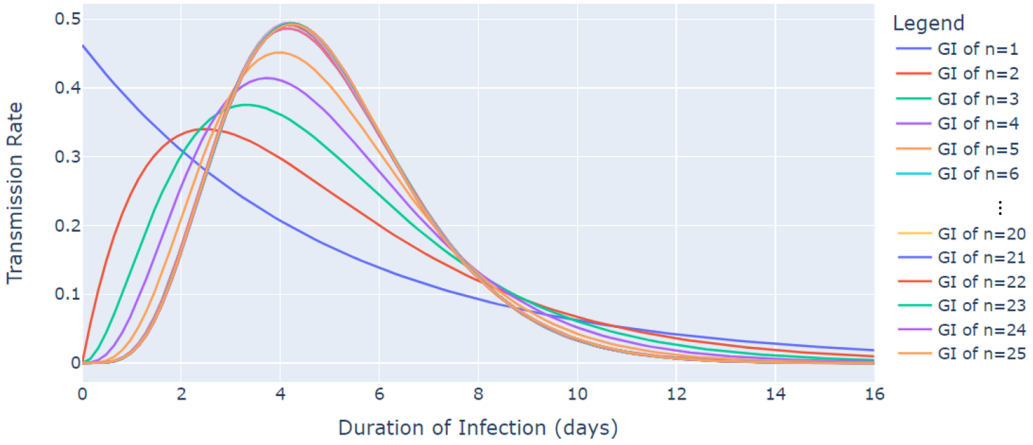

Figure 8: Expected Transmission Curve $\beta(\tau)$ of a Gamma Distributed Generation Interval

first five of these compartments. We noticed that many of the initial compartments in this set had greater weights for the sixth and seventh compartment. Upon closer examination, we realized that the sample shape of the simulated distribution was 6.246 and surmised that the weights and shape parameter were linked. Based on this and the weights returned, we found that the fractional part of the shape parameter dictated the compartments weights. For example, since the shape of this distribution was 6.246, we found that the sixth compartment should have had an initial estimate of 1-0.246=0.754 while the seventh compartment should have a weight of roughly 0.246.

Another similarity we noticed was that the value of the holding period was near 1.3 for compartments 6 through 10. We found that 1.3 was close to mean divided by the variance of the gamma distributed generation interval data. Based on all of these findings, we modified the estimates to reflect the weights link to the sample shape and as well as the holding period relation to $\frac{mean}{variance}$. To determine the best initial estimates for the log-normal and Weibull transmission results, we first examined the data with the same initial estimates the gamma distribution used. Then we compared these results with the log-normal and Weibull distributions using the preferred initial estimates for the gamma distribution.

While examining the expected infectious curve of the log-normal distribution, we recognized that there were two main overlapping curves. The first set was for $n = 7, ..., 13$ while the second was for

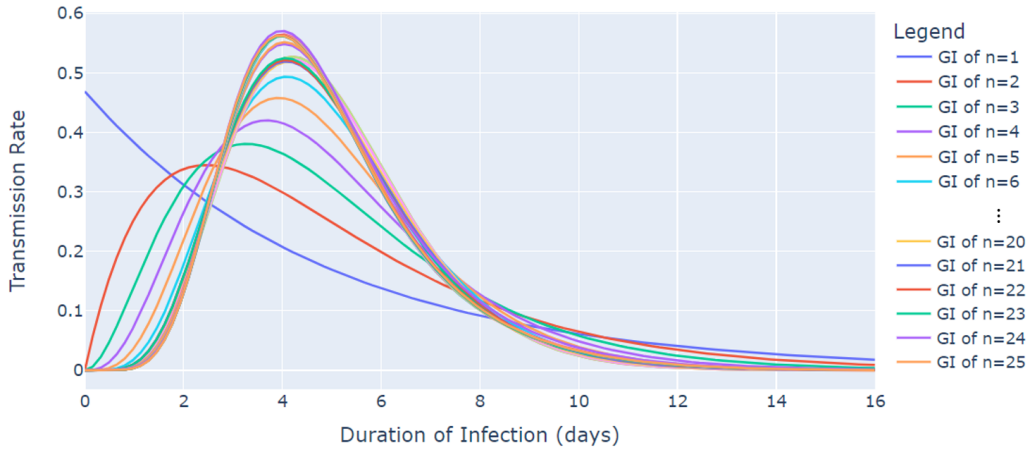

Figure 9: Expected Transmission Curve $\beta(\tau)$ of a Log-normal Distribution

$n = 14, ..., 25$. The holding period of the first set was typically close to 1.48 and the majority of the weight (≈ 0.9) resided in the seventh compartment. There was an average holding period of approximately 1.8 within the second set of compartments (n=14,...,25). The majority of the weight shifted between the eighth and ninth compartments depending on the compartment size. As the holding periods primarily fell between 1.2 and 2 for the majority of the compartments, we found that a decent estimate would be the normal mean μ_n of the distribution which in this case was 1.512. As the weight assigned to the compartments varied drastically by compartment size, we kept the weight estimates as $w_j = \frac{1}{n} \; \forall \; j \in 1, .., n$. Next, we examined the expected transmission curve of the Weibull distributed generation interval.

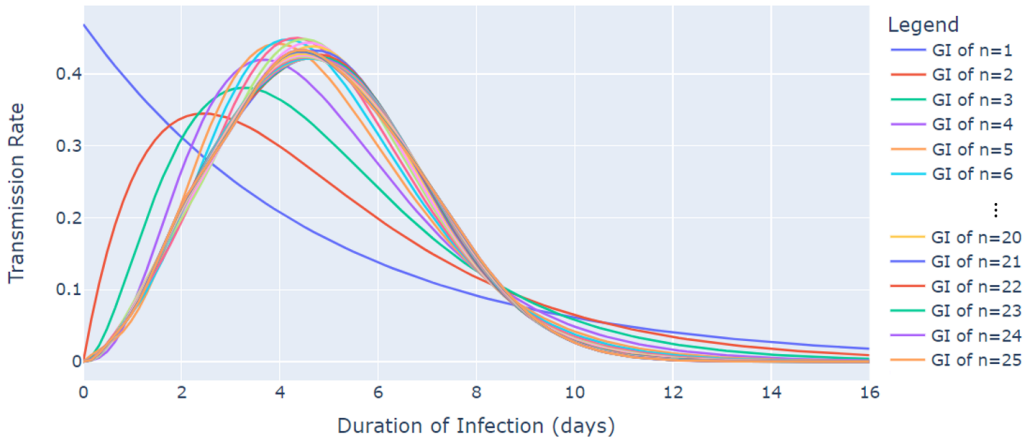

Figure 10: Expected Transmission Curve $\beta(\tau)$ of a Weibull Distribution

Unlike the curve implementing the gamma and log-normal distributed data sets, the expected infectious curve of the Weibull distributed generation interval did not have as clear defined curves after a specific n value. For compartments sizes smaller than $n = 13$, the weight in the final compartment $w_n \geq 0.5$. The first n =1,...,8 compartment sizes each had the final weight w_n of more than 0.8 out of 1. We also noticed that the average holding period value was about 2. For this reason, we changed the initial estimate of the holding period to be the standard deviation of the distribution which equalled 2.055. We tried changing the weight estimate to place 0.5 in the final compartment and split the rest evenly; however, we found this was not the best estimate. We ultimately returned the initial weight estimates to $w_j = \frac{1}{n} \; \forall \; j \in 1, .., n$.

After finding the best initial estimates from examining the results, we then tested the gamma distribution estimates for the log-normal and Weibull distributions. We did this because the transmission rates would ultimately follow a mixture of gamma distributions regardless of the given distribution of the generation interval. The table below shows the difference in objective values based on the minimized solution from the python algorithms explained in Section 3.4.

n	Log-normal Objective Results			Weibull Objective Results		
	Log-normal	Gamma	Difference	Weibull	Gamma	Difference
7	10024.672	10024.628	0.044	10635.693	10635.694	-0.00032
8	10024.672	10024.661	0.010	10622.269	10622.269	0.000014
9	10020.601	10020.597	0.0042	10614.944	10614.945	-0.0006
10	10012.403	10012.402	0.001	10610.319	10610.320	-0.0003
11	10004.828	10004.881	-0.054	10607.422	10607.422	0.0002
12	9998.392	9999.8446	-1.452	10605.691	10605.694	-0.002
13	9989.417	9997.242	-7.824	10604.711	10604.710	0.002

Table 5: Objective Values based on Initial Estimate Parameters

The table above shows the returned objective value of the model based on initial estimates. The "Difference" columns calculate the difference between the predicted log-normal or Weibull distribution estimate and the gamma distribution estimates. The python used a minimize function. Therefore, the smaller the objective value, the better. These values were rounded to the nearest ten thousandths place. As such, the difference may not be the exact same.

We noticed that the log-normal distribution performed better for compartments immediately following the minimum accepted compartment size. More specifically, compartments seven through ten performed better with initial gamma estimates. Once the compartment size exceed ten, the log-normal estimate performed better. This indicated to us that the log-normal shift between performance could be a good indicator of a sufficiently large compartment size. We chose to use the gamma estimate for the log-normal distribution.

Although we would expect the gamma assumed estimates to provide better results for both distributions, their objective value was typically worse for the Weibull distribution. For this reason, we opted for the Weibull estimate explained previously.

4.1.2 Compartment Size Selection

With the preferred initial estimates implemented in the code for each of the three distributions, we then examined the expected infectious curve for compartments $n = 7, ..., 25$ for each of the distributions.

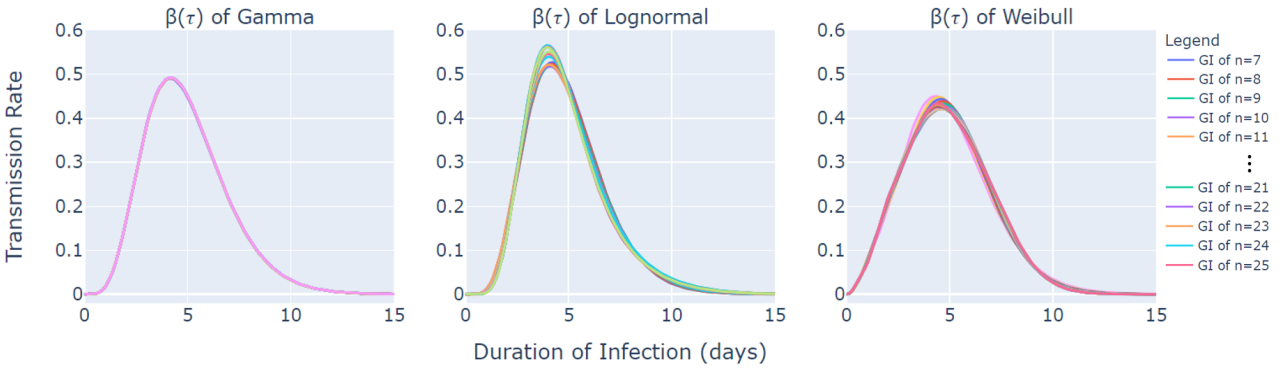

Figure 11: Expected Infectious Curve $\beta(\tau)$ for each Distribution given Preferred Estimates

Based on the initial estimates for the gamma, log-normal, and Weibull distributions, we knew that the compartment size had to be sufficiently large. For the gamma and log-normal distributions, this meant they required a minimum compartment size of the shape parameter. The gamma and log-normal shape parameters were 6.34 and 6.19 respectively. For this reason, they needed a minimum compartment size of 7. As the Weibull distribution was not beholden to the same weight estimates, the Weibull could technically have had any compartment size. However, based on Figure 10, a minimum compartment size of $n = 5$ was more suitable.

In addition to ensuring a large enough compartment size, we also needed to confirm that the compartment size was not too large as this could also cause problems with finding reasonable results.

For example, if we found that the majority of the weight was in the fifth and sixth compartment regardless of compartment size, then having a compartment size of $n = 100$ would be unnecessary as compartments seven through 100 will be near or at a transmission rate of 0. This could lead to incorrectly labelling someone as infected when in reality they should be in the recovered or resistant compartments of the SI_nR and SI_nRS models respectively. We ultimately chose a compartment size of $n = 10$ as this was larger than all three values and we believed it to be sufficient for the distributions' respective compartment sizes.

4.2 Transmission Rates

Given a compartment size of $n = 10$, we then found the respective transmission weights w_j^* and holding period λ^* for each distribution. They are as follow:

Distribution	**Transmission weights w_j^***	**Holding Period λ^***
Gamma	0, 0, 0, 0, 0, 0.507, 0.459, 0.033, 0, 0	1.312
Log-normal	0, 0, 0, 0, 0, 0, 0.894, 0, 0, 0.106	1.492
Weibull	0, 0, 0.026, 0, 0.101, 0.178, 0, 0, 0, 0.695	1.754

Table 6: Transmission Weights and Holding Period for $n = 10$

The weights and holding period values returned were similar to the initial values used to estimate the parameters for each distribution. As previously mentioned, the shape of the gamma distribution was 6.246. As such, the initial estimate was 0.754 in compartment six and 0.246 in compartment in 7. The values 0.507 and 0.459 were close to these initial estimates. Furthermore, the holding period we set as the initial estimate was equal to 1.255 given the generation interval data of the gamma distribution. This value was close to the minimized holding period of 1.312. The log-normal distribution placed the majority of its weight in the seventh compartment with the estimated holding period of 1.492. The shape of the log-normal distribution was 6.192. As such, we estimated 0.192 to be in the sixth compartment and 0.807 to be in the seventh compartment. The Weibull distribution also followed the trends we saw in the previous section as the final compartment contained most of the weight and the holding period was near 2.

With all of the parameters calculated, we used Equation 3.11 to calculate the transmission rates.

Distribution	**Transmission Rates β_j^***	**Objective Value**
Gamma	0, 0, 0, 0, 0, 1.53, 1.386, 0.101, 0, 0	10275.067
Log-normal	0, 0, 0, 0, 0, 0, 3.068, 0, 0, 0.363	10012.402
Weibull	0, 0, 0.105, 0, 0.409, 0.718, 0, 0, 0, 2.803	10610.319

Table 7: Transmission Rates for $n = 10$

These transmission rates were used for the rest of the results. An unexpected finding was that the gamma distribution did not have the lowest objective value. As the transmission function was a combination of gamma distributions, we expected the gamma distribution to have the best objective value. Another interesting observation was that the Weibull objective was larger than the gamma and log-normal distributions. This is unsurprising as it did not fit well to the gamma initial estimates. As these values are relatively close, we believed these transmission rates to be acceptable.

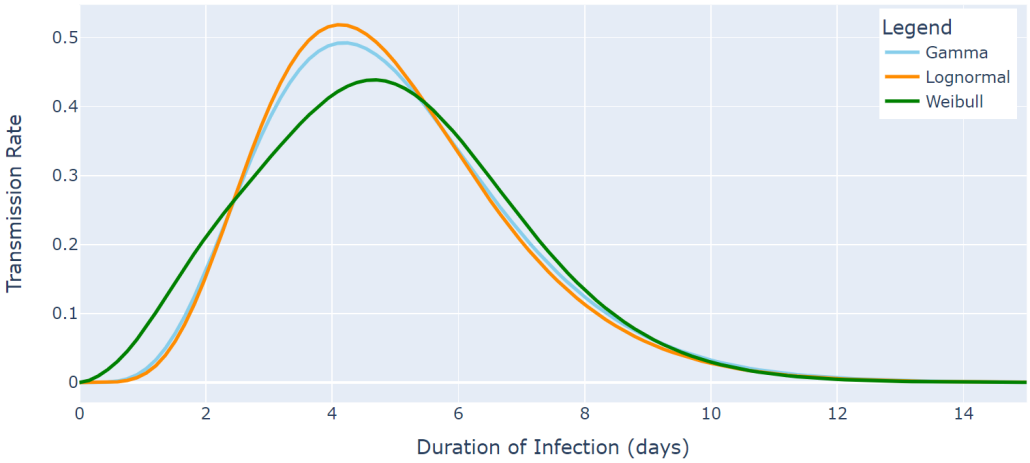

Figure 12: Expected Transmission Curve $\beta(\tau)$ of all Distributions for $n = 10$

The figure above illustrates the expected transmission curve $\beta(\tau)$ for each of the distributions with using their respective transmission rates shown in Table 7. These results seem reasonable as they are similar to the generation intervals shown in Figure 6.

4.3 Compartmental Models

From Section 3.1, we selected a basic reproduction number of 2.3 and determined that all three distributions should be based on Challen's 2020 study [10] in which he found a population mean and standard deviation of a gamma distribution to be 4.9 and 2 respectively. We then found that in 2020, the UK had a recorded population of 67,886,011 [40] which we used as our population size N. We made the assumption that no one in the population was immune from the disease. Additionally, we assumed that only one person was infected initially and began in compartment I_1. The constant parameters used for the SI_nR and SI_nRS compartmental models are shown in the table below.

Parameter	Definition	Value
R_0	Basic reproduction number	2.3
τ	Generation Interval $[\tau_1, ..., \tau_{5000}]$	$\mu = 4.9$ and $\sigma = 2$
N	Total population	67,886,011
$S(0)$	Number of susceptible individuals	67,886,010
$I_1(0)$	Number of infected people in compartment 1	1
$I_2(0)$ to $I_n(0)$	Number of infected people in compartments 2 through n	0
$R(0)$	Number of recovered individuals	0

Table 8: Parameter Selection for SI_nR and SI_nRS models.

4.3.1 SI_nR Model Results

As the SI_nR model assumes life long immunity, we set the recovery rate h to 0. We also set the contact rate to 1, meaning that we assumed no non-pharmaceutical interventions or vaccinations occurred. With all of the parameters set for the SI_nR model, we modelled each of the three generation interval distributions.

In the figure below, the results were shown for each of the three distributions. The gamma distributed generation interval data, displayed in solid lines, predicted peak prevalence at 30% of the population within 103 days. The log-normal data provided similar results with peak prevalence at 102 days with 27% of the population. The Weibull model, show cased as small dotted lines, provided

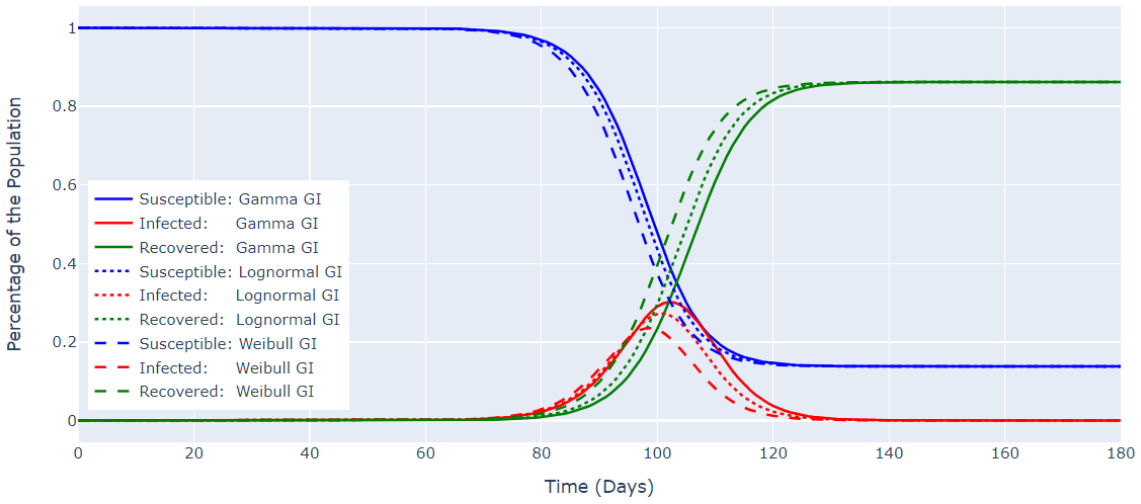

Figure 13: $SI_{10}R$ Model Results for Contact Rate $c = 1$

the lowest overall peak for prevalence with only 23% at 93 days. Each of these distributions can be viewed separately in Appendix A.

Although the log-normal distribution had the largest transmission rate, the gamma distribution predicted the greatest prevalence. To better compare the predicted prevalence cases, we isolated the sum of the infected compartments.

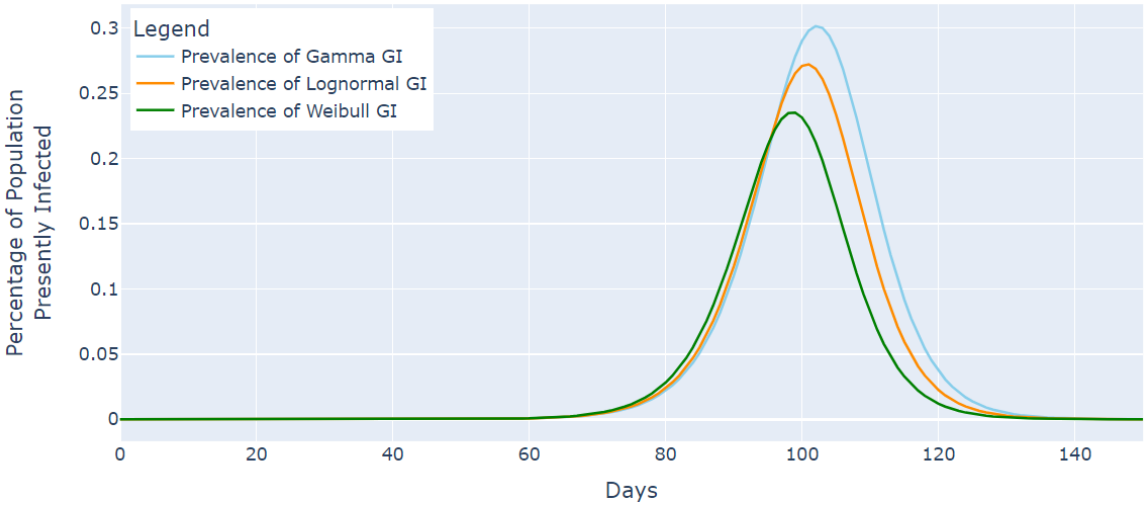

Figure 14: Expected Prevalence for each Generation Interval Distribution for Contact Rate $c = 1$

The figure above more clearly displayed how each of the distributions compared to one another when applied to the SI_nR compartmental model. From this figure, we more clearly saw that the Weibull distribution initially increased exponentially at a faster rate than the log-normal and gamma distributions. The Weibull distribution also peaked about a week and a half prior to the log-normal and gamma distributions. Although the log-normal distribution provided a slightly higher peak value for transmission (see Figure 12), the gamma distributed generation interval provided the greatest prevalence case expectation.

Even though the distributions all stemmed from the same population mean and standard deviation and had similar sample means and standard deviations, their expected COVID-19 prevalence cases varied. This was in part due to the variation in the other order moments rather than just the mean and variance. As time progressed, these subtle differences became more evident through exponential rise. We then compared these predicted prevalence cases to the real UK prevalence cases in the United Kingdom.

The UK had it's first confirmed case of COVID-19 on January 29, 2020 [6]. On March 18th schools

shut down [6]. Five days later, on March 23, 2020, Prime Minister Boris Johnson announced UK's first official lockdown in which schools, pubs, gyms, and more were officially closed [6]. The figure below displays the SI_nR model results with the first infection on 29th of January, 2020 and follows two weeks past the lockdown in April 2020.

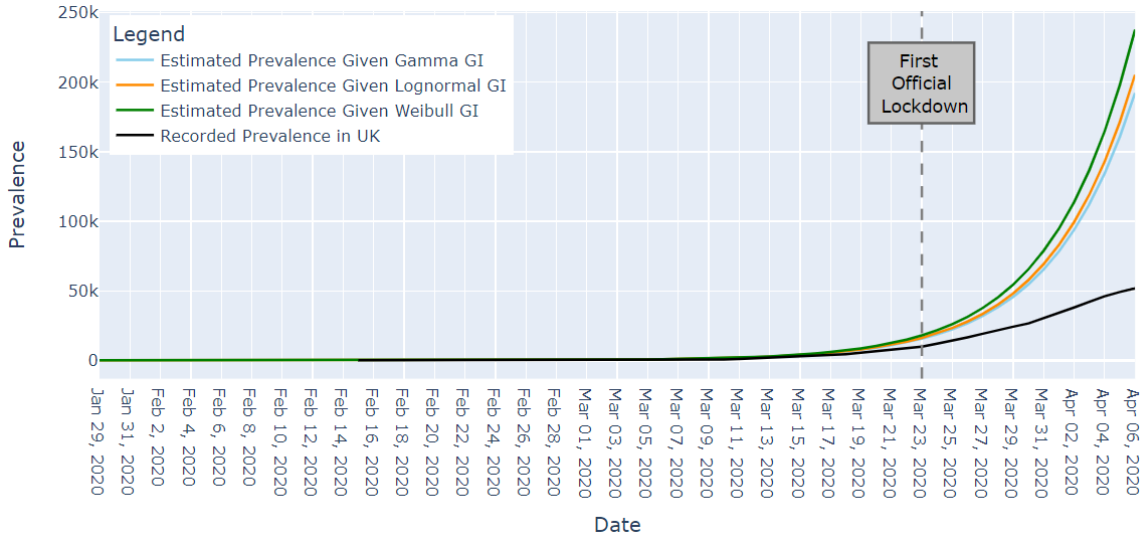

Figure 15: Expected vs Real Prevalence for Contact Rate $c = 1$

The recorded prevalence data, taken from Worldometer, had its first official account of COVID-19 on February 15 with its acknowledgement of 33 prevalence cases [41]. Worldometer calculated prevalence by current cases minus deaths and recovered cases. The blue, orange, and green curves are the predicted prevalence cases given a generation interval that had gamma, log-normal, and Weibull distributions respectively.

Before large non-pharmaceutical interventions were imposed on the population, the prevalence cases appeared to closely align. At the point of lockdown, prevalence in the UK was 9,842. The gamma, log-normal, and Weibull distributions estimated the prevalence to be 14,698, 15,295, and 16,993 respectively. Upon closer inspection in the figure below, we can see more clearly these estimates.

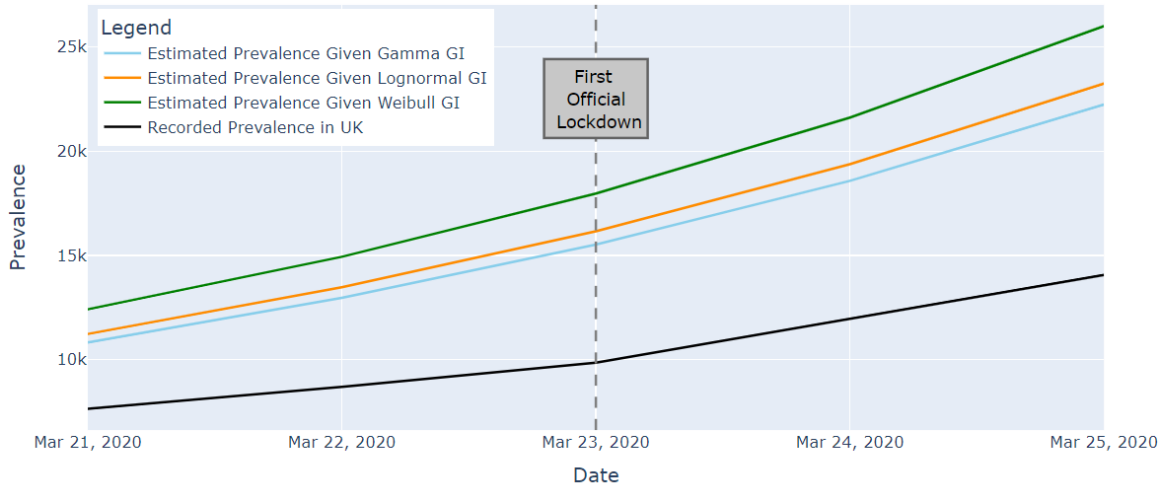

Figure 16: Magnified Expected vs Real Prevalence for Contact Rate $c = 1$

There were two primary explanations for the elevated predicted prevalence cases. The first point considered was asymptomatic cases. Many scientists predicted that, although many countries reported the incidence and prevalence cases, these estimates were low. Although the UK had 3,282 [41] confirmed cases by March 16, 2020, it was believed by many that at least 10,000 individuals had already

been infected with COVID-19 [6]. Because COVID-19 had both symptomatic and asymptomatic individuals, analysts speculated that the number of cases was higher due to asymptomatic people not testing.

Additionally, this model assumed no non-pharmaceutical interventions occurred. However, this was not the case. Before the official lockdown in the UK, individuals who tested positive for COVID-19 in the first few weeks and needed medical care were quarantined in the hospital. These limited quarantines provided additional insight as to why the predicted prevalence cases for each of the distributions were slightly elevated.

To understand how contact changed during Covid, we experimented with decreased values of contact rate c. We changed the contact rate from $c = 1$ to $c = 0.97$ and found that the predicted cases and recorded cases were much closer. The gamma, log-normal, and Weibull distributions predicted 10,245, 10,622, 11,602 cases as compared to the recorded prevalence cases of 9,842. These case

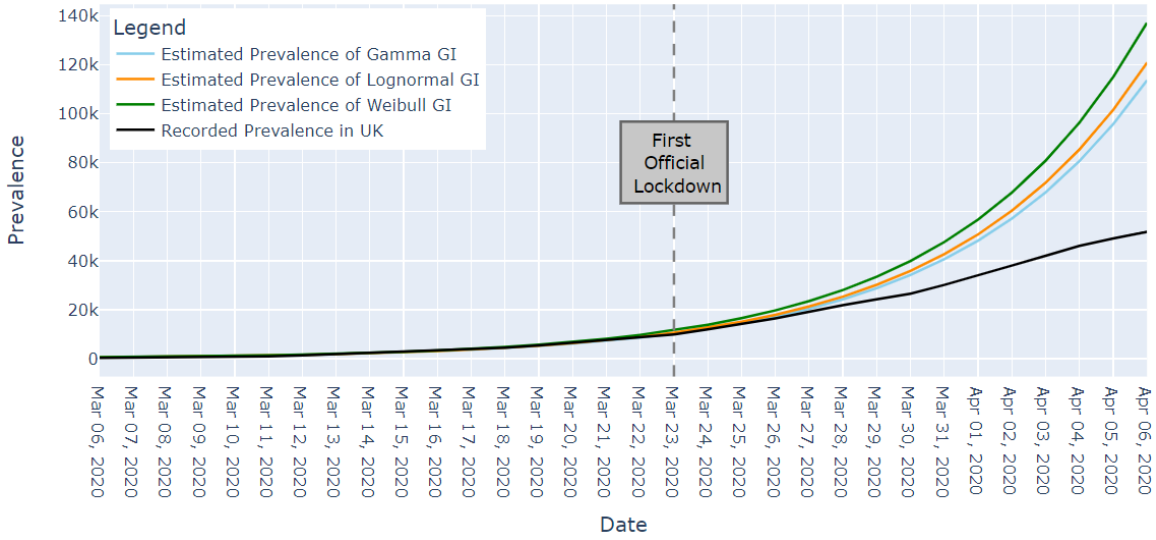

Figure 17: Expected vs Real Prevalence for Contact Rate $c = 0.97$

predictions were much closer to the actual cases in the UK and further supported the idea that minimal quarantines had a small impact on the population. Based on the figure above, the school lockdown, implemented on the 18th of March, followed by the official lockdown of non-essential travel was an important moment in the decrease of prevalence cases. The graph suggests that if no major interventions were imposed, then many more people would have been infected. With the contact rate decreased to 0.97, we reran the SI_nR model.

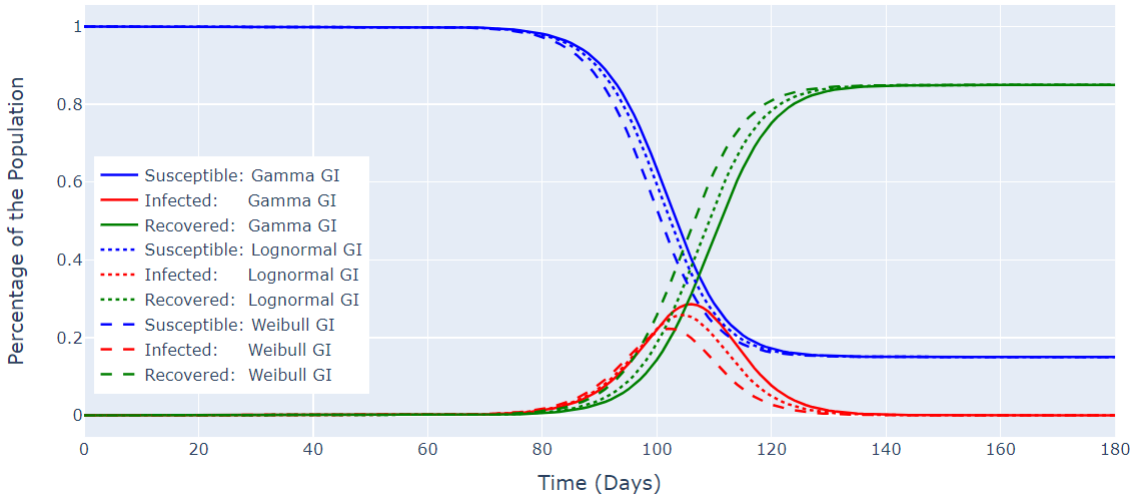

Figure 18: $SI_{10}R$ Model Results for Contact Rate $c = 0.97$

By accounting for the minimal non-pharmaceutical interventions present at the beginning of the pandemic in the United Kingdom, we found a slight decrease in peak prevalence for the gamma, log-normal and Weibull distributions. Peak prevalence occurred at 107, 106, and 104 days after the introduction of the first infected individual. Their maximum estimations included 28%, 25.3%, and 21.8% of the population with prevalence cases. To further understand the implications of non-pharmaceutical intervention, we varied the contact rate for the gamma distribution.

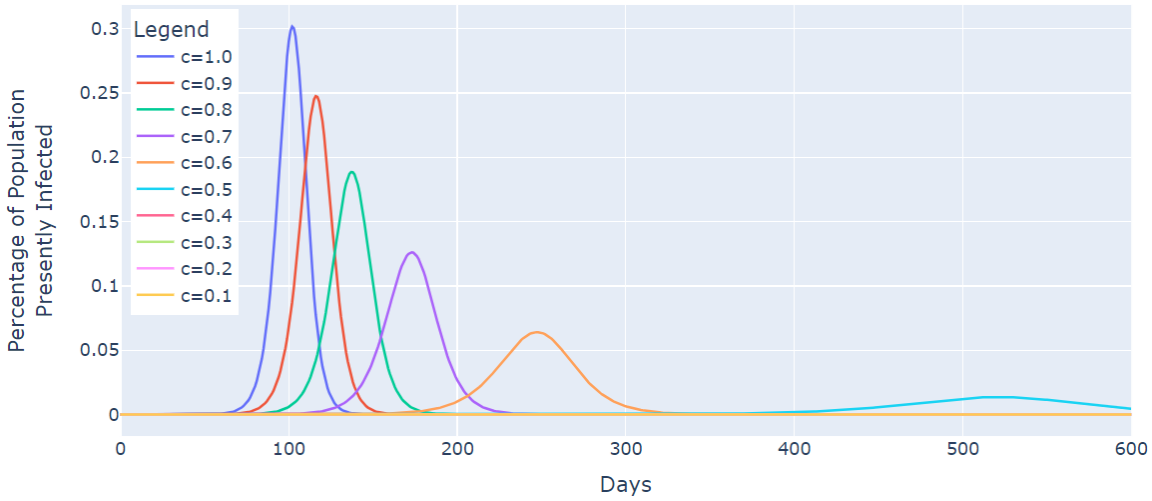

Figure 19: SI_nR Model Results of Varying Contact Rate c of the Gamma Distribution

By modifying c values, we saw how the prevalence cases decreased overtime. One study [35] grouped non-pharmaceutical interventions into five categories ranging from high impact to low impact on contact rates. It found that suspension of public transportation and cancelling of large events had a high impact on decreasing contact while closure of schools and day cares had a moderately low impact on contact rates [35]. Base on additional information, one could set estimates for each impact and visualize the effects. For example, this study found that working from home had a moderate impact on contact rate c [35]. As such we could assume that a "moderate impact" carried an estimated contact rate of 0.7. If this were true, we would then find that this mandate would have had to stay in place for over 200 days (purple line in figure above) to decrease infected population to an extremely low value.

Herein lies the importance of epidemiological predictions for policy officials. As the model highlighted estimated prevalence, policy officials had to balance the safety of people as well as their continued well being. Different non-pharmaceutical interventions and their duration affected physical mental health, the economy, and so much more. The need to provide effective health precautions while balancing the needs of the population was an extremely tricky task. These models aided some policy makers as to what interventions and when to implement. In addition to these short term challenges policy officials had to face, we also examined the long-term consequences.

4.3.2 SI_nRS Model Results

The SI_nRS model assumed that waning immunity existed. We assumed that natural immunity lasted 4 months, or 122 days. As such, we set $h = \frac{1}{122}$. We also wanted to examine how the predicted prevalence would change overtime given no non-pharmaceutical interventions nor vaccinations. This model also assumes that no other variants were created. Therefore, we initially examined the SI_nRS model with $c = 1$. These distribution model results can be found separately in Appendix B.

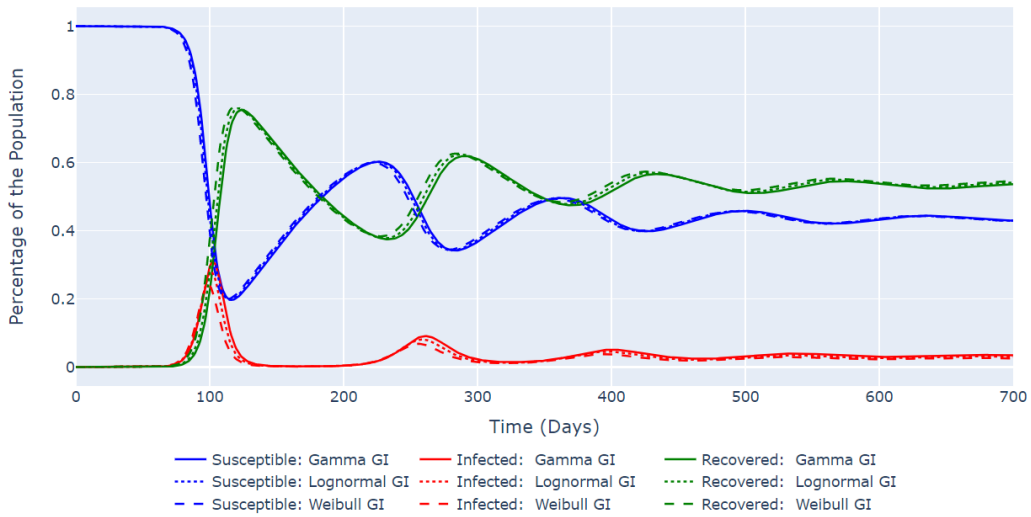

Figure 20: $SI_{10}RS$ Model Results for Contact Rate $c = 1$ and Waning Immunity Rate $h = \frac{1}{122}$

Given these assumptions and initial inputs, the model predicted that COVID-19 would eventually end in the UK (or at least become very minimal). As it is hard to clearly see the interaction between predicted prevalence cases, we created the figure below to isolate this data.

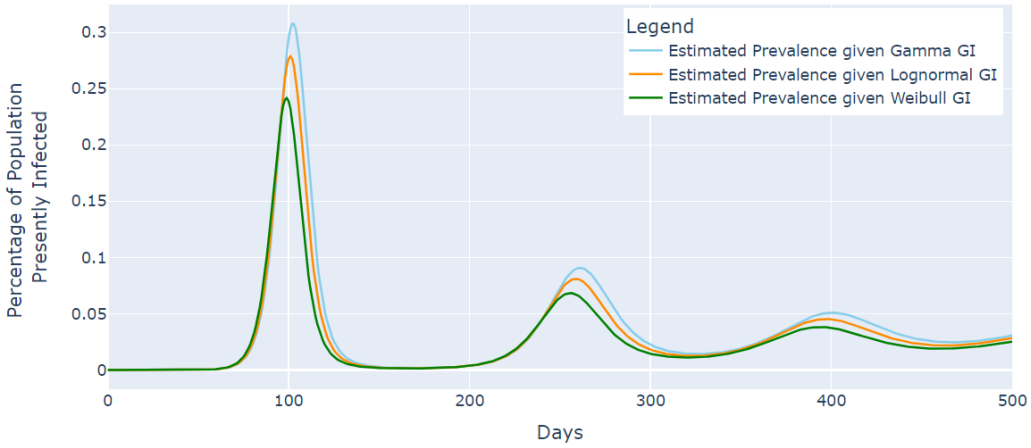

Figure 21: Expected Prevalence for each Generation Interval Distribution

We saw that, over time, the prevalence peaks decreased. The first, second, and third prevalence peaks of the gamma distribution were at 30% , 9%, and 5% of the population. These peaks highlighted the waning immunity rate and how it affected the population overtime.

5 Discussion and Conclusion

5.1 Comparing the Gamma, Log-normal, and Weibull Distributions

There were a few results that peaked our interest with regards to the distributions. One interesting result we found was that the log-normal initial estimates provided better results depending on compartment size. More specifically, the gamma estimates provided better results for the log-normal distribution for all values of $n < 11$. Once $n \geq 11$, the log-normal initial estimates typically provided better results. The relationship between the log-normal and gamma initial estimates could have been an indicator of acceptable compartment sizes. The relationship between these two distributions could, in future work, provide greater understanding of what the reasonable range of compartment sizes are.

Another interesting result we found was the objective value differences when juxtaposed with the model itself. The figure below highlights the objective value results for each of the distributions for $n \in 7, ..., 10$.

n	**Gamma Objective**	**Log-normal Objective**	**Weibull Objective**
7	10275.071	10024.628	10635.693
8	10275.022	10024.661	10622.269
9	10275.066	10020.597	10614.944
10	10275.067	10012.402	10610.319

Table 9: Objective Values of Distributions given Varying Compartment Sizes

We found that the log-normal generation interval distribution typically had the lowest objective value. Although we were surprised that the log-normal distribution had the best objective value, the gamma distribution was a close second. The Weibull distribution however, was relatively higher than the other two. Even though the log-normal distribution provided the best objective value, the gamma distributed generation interval data provided the most realistic SI_nR model as it's predicted prevalence cases were closest to the real prevalence data. The log-normal data provided the second most realistic data while the Weibull distribution was least realistic.

5.2 Summary of Work

In this project, we studied generation intervals and their impact on the SI_nR and SI_nRS models. Based on previous literature, we used estimated values for key epidemiological parameter in conjunction with simulated generation interval data sets to implement two compartmental models. These models provided interesting insights.

We first found that, although the sample means and standard deviations were similar for the different generation interval data sets, the exponential rise introduced through model implementation created varying results. Another interesting aspect we found from our research was that the gamma provided the most realistic data followed closely by the log-normal distribution. With the models implemented, we were also able to test the contact rates. These contact rates gave a peak into the challenges that policy officials needed to make as the balance between public safety and well-being was tedious. This study provided us with the opportunity to examine COVID-19 in a way many studies have yet to use. Because of this, there are a lot of interesting aspects to continue to study within this model.

5.3 Future Work

This project lead to many interesting results and potential extensions. Given this potential dynamic between compartment size and the gamma and log-normal distributions, it would be fascinating to investigate this concept in depth. If this dynamic proved to be useful, we could quickly solve for reasonable compartment sizes and select a model from one or all options.

In the future, we also hope to examine the other order moments to see if there is a clear relationship between the orders and the distributions. The third and fourth order moments of skewness and

kurtosis could explain the limitations of the Weibull distribution we found and provide interesting findings not previously analyzed. Once we have an even clearer image of the distributions and how they relate, we could potentially extend this project to other distributions. In addition to examining other distributions and ordered moments, we would prefer to extend the model to include vaccinations. By examining the SVI_nR and SVI_nRS models, we could extend the model past the United Kingdom's first lockdown. This would allow the model to more properly account for the reality of 2021 and 2022.

Code: All code related to this project can be found at the GitHub link: `https://github.com/annette-bell/SInR-Covid-Dissertation`. In addition to the code, we also created a dash app to test different values. In addition to a video of the dash application in the GitHub link.

References

[1] SIR models of epidemics.

[2] Reproduction number (R) and growth rate (r) of the covid-19 epidemic in the uk: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation. *The Royal Society*, 2020.

[3] Covid-19: Epidemiology, virology and clinical features, May 2022.

[4] B. M. Althouse, E. A. Wenger, J. C. Miller, S. V. Scarpino, A. Allard, L. Hébert-Dufresne, and H. Hu. Stochasticity and heterogeneity in the transmission dynamics of sars-cov-2, 2020.

[5] R. M. Anderson. Discussion: The kermack-mckendrick epidemic threshold theorem. *Bulletin of Mathematical Biology*, 53(1):3–32, 1991.

[6] E. Aspinall. COVID-19 timeline, Mar 2022.

[7] T. S. Brett and P. Rohani. COVID-19 herd immunity strategies: walking an elusive and dangerous tightrope. *medRxiv*, May 2020.

[8] A. W. Byrne, D. McEvoy, A. B. Collins, K. Hunt, M. Casey, A. Barber, F. Butler, J. Griffin, E. A. Lane, C. McAloon, K. O'Brien, P. Wall, K. A. Walsh, and S. J. More. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ Open*, 10(8):e039856, Aug. 2020.

[9] Centers for Diseases Control and Prevention. CDC museum covid-19 timeline, 2022.

[10] R. Challen, E. Brooks-Pollock, K. Tsaneva-Atanasova, and L. Danon. Meta-analysis of the severe acute respiratory syndrome coronavirus 2 serial intervals and the impact of parameter uncertainty on the coronavirus disease 2019 reproduction number. *Stat. Methods Med. Res.*, page 9622802211065159, Dec. 2021.

[11] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio r0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, Jun 1990.

[12] S. E. Eikenberry, M. Mancuso, E. Iboi, T. Phan, K. Eikenberry, Y. Kuang, E. Kostelich, and A. B. Gumel. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infectious Disease Modelling*, 5:293–308, 2020.

[13] N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, and et al. Report 9 - impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand, Mar 2020.

[14] L. Ferretti, A. Ledda, C. Wymant, L. Zhao, V. Ledda, L. Abeler-Dörner, M. Kendall, A. Nurtay, H.-Y. Cheng, T.-C. Ng, H.-H. Lin, R. Hinch, J. Masel, A. M. Kilpatrick, and C. Fraser. The timing of COVID-19 transmission. Sept. 2020.

[15] E. C. for Disease Prevention and Control. Coronavirus disease 2019 (covid-19) in the eu/eea and the uk - ninth update, Apr 2020.

[16] T. Ganyani, C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, march 2020. *Euro Surveill.*, 25(17), Apr. 2020.

[17] Y. Goldberg, M. Mandel, Y. M. Bar-On, O. Bodenheimer, L. S. Freedman, N. Ash, S. Alroy-Preis, A. Huppert, and R. Milo. Protection and waning of natural and hybrid immunity to SARS-CoV-2. *New England Journal of Medicine*, 386(23):2201–2212, June 2022.

[18] W. S. Hart, S. Abbott, A. Endo, J. Hellewell, E. Miller, N. Andrews, P. K. Maini, S. Funk, and R. N. Thompson. Inference of the sars-cov-2 generation time using uk household data. *eLife*, 11:e70767, Feb 2022.

[19] W. S. Hart, P. K. Maini, C. A. Yates, and R. N. Thompson. A theoretical framework for transitioning from patient-level to population-scale epidemiological dynamics: influenza a as a case study. *J. R. Soc. Interface*, 17(166):20200230, May 2020.

[20] J. A. P. Heesterbeek. A brief history of R_0 and a recipe for its calculation. *Acta Biotheoretica*, 50(3):189–204, Sep 2002.

[21] M. Jit, T. Jombart, E. S. Nightingale, A. Endo, S. Abbott, and W. J. E. and. Estimating number of cases and spread of coronavirus disease (COVID-19) using critical care admissions, united kingdom, february to march 2020. *Eurosurveillance*, 25(18), May 2020.

[22] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A Math. Phys. Sci.*, 115(772):700–721, Aug. 1927.

[23] O. Krylova. Predicting epidemiological transitions in infectious disease dynamics: Smallpox in historic london (1664-1930). Oct. 2011.

[24] M. Li, K. Liu, Y. Song, M. Wang, and J. Wu. Serial interval and generation interval for imported and local infectors, respectively, estimated using reported contact-tracing data of COVID-19 in china. *Front. Public Health*, 8:577431, 2020.

[25] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13):1199–1207, Mar. 2020.

[26] K. Linka, M. Peirlinck, and E. Kuhl. The reproduction number of COVID-19 and its correlation with public health interventions. *medRxiv*, July 2020.

[27] E. Lippiello, G. Petrillo, and L. de Arcangelis. Estimating the generation interval from the incidence rate, the optimal quarantine duration and the efficiency of fast switching periodic protocols for COVID-19. *Sci. Rep.*, 12(1):4623, Mar. 2022.

[28] M. Lonergan and J. D. Chalmers. Estimates of the ongoing need for social distancing and control measures post-"lockdown" from trajectories of COVID-19 cases and mortality. *European Respiratory Journal*, 56(1):2001483, June 2020.

[29] J. Lourenço, R. Paton, C. Thompson, P. Klenerman, and S. Gupta. Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. Mar. 2020.

[30] J. S. Marks. Epidemiology, public health, and public policy. *Prev. Chronic Dis.*, 6(4):A134, Oct. 2009.

[31] E. Nepomuceno, D. Resende, and M. Lacerda. A survey of the individual-based model applied in biomedical and epidemiology. 02 2019.

[32] T. W. Ng, G. Turinici, and A. Danchin. A double epidemic model for the SARS propagation. *BMC Infectious Diseases*, 3(1), Sept. 2003.

[33] E. Okyere, J. D.-G. Ankamah, A. K. Hunkpe, and D. Mensah. Deterministic epidemic models for ebola infection with time-dependent controls. *Discrete Dynamics in Nature and Society*, 2020:1–12, July 2020.

[34] S. W. Park, K. Sun, D. Champredon, M. Li, B. M. Bolker, D. J. D. Earn, J. S. Weitz, B. T. Grenfell, and J. Dushoff. Forward-looking serial intervals correctly link epidemic growth to reproduction numbers. *Proceedings of the National Academy of Sciences*, 118(2), Dec. 2020.

[35] D. F. Patiño-Lugo, M. Velez, P. V. Salazar, C. Y. Vera-Giraldo, V. Ve

lez, I. C. Mari

n, P. A. Rami

rez, S. P. Quintero, E. C. Martínez, D. A. P. Higuita, and G. N. H. Herrera. Non-pharmaceutical interventions for containment, mitigation and suppression of COVID-19 infection. *Colombia Medica*, pages 1–25, May 2020.

[36] B. Ridenhour, J. M. Kowalik, and D. K. Shay. Unraveling R_0: Considerations for public health applications. *Am. J. Public Health*, 108(S6):S445–S454, Dec. 2018.

[37] R. ud Din and E. A. Algehyne. Mathematical analysis of COVID-19 by using SIR model with convex incidence rate. *Results in Physics*, 23:103970, Apr. 2021.

[38] T. Wang, Y. Wu, J. Y.-N. Lau, Y. Yu, L. Liu, J. Li, K. Zhang, W. Tong, and B. Jiang. A four-compartment model for the COVID-19 infection—implications on infection kinetics, control measures, and lockdown exit strategies. *Precision Clinical Medicine*, 3(2):104–112, May 2020.

[39] Worldometer. Coronavirus worldwide graphs, 2022.

[40] Worldometer. U.k. population, 2022.

[41] Worldometer. United kingdom, 2022.

[42] C. Xu, Y. Dong, X. Yu, H. Wang, L. Tsamlag, S. Zhang, R. Chang, Z. Wang, Y. Yu, R. Long, Y. Wang, G. Xu, T. Shen, S. Wang, X. Zhang, H. Wang, and Y. Cai. Estimation of reproduction numbers of COVID-19 in typical countries and epidemic trends under different prevention and control scenarios. *Frontiers of Medicine*, 14(5):613–622, May 2020.

[43] J. Zhang, M. Litvinova, W. Wang, Y. Wang, X. Deng, X. Chen, M. Li, W. Zheng, L. Yi, X. Chen, Q. Wu, Y. Liang, X. Wang, J. Yang, K. Sun, I. M. Longini, M. E. Halloran, P. Wu, B. J. Cowling, S. Merler, C. Viboud, A. Vespignani, M. Ajelli, and H. Yu. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside hubei province, china: a descriptive and modelling study. *The Lancet Infectious Diseases*, 20(7):793–802, July 2020.

[44] S. Zhao, B. Tang, S. S. Musa, S. Ma, J. Zhang, M. Zeng, Q. Yun, W. Guo, Y. Zheng, Z. Yang, Z. Peng, M. K. Chong, M. Javanbakht, D. He, and M. H. Wang. Estimating the generation interval and inferring the latent period of covid-19 from the contact tracing data. *Epidemics*, 36:100482, 2021.

[45] A. Zlojutro, D. Rey, and L. Gardner. A decision-support framework to optimize border control for global outbreak mitigation. *Scientific Reports*, 9(1):2216, Feb 2019.

Appendices

A $SI_{10}R$ Compartmental Model Results for Gamma, Weibull, and Log-normal

Below are each of the results for the $SI_{10}R$ model in their own figure. To compare each of these directly, see Figure 13.

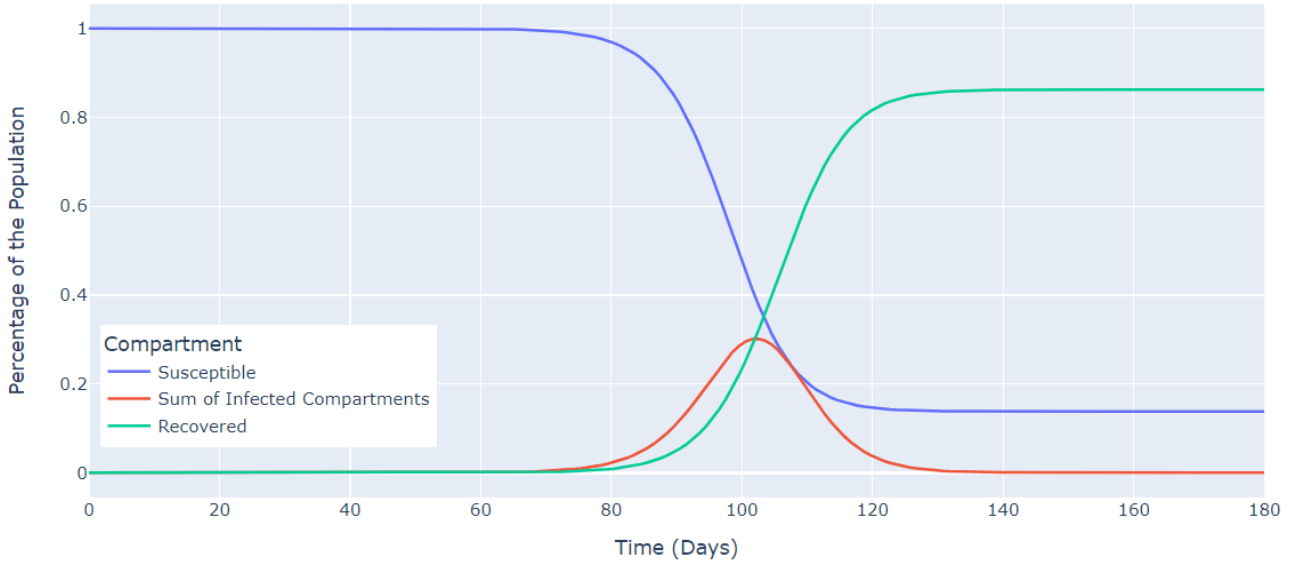

Figure 22: $SI_{10}R$ Model of Gamma Distributed Generation Interval

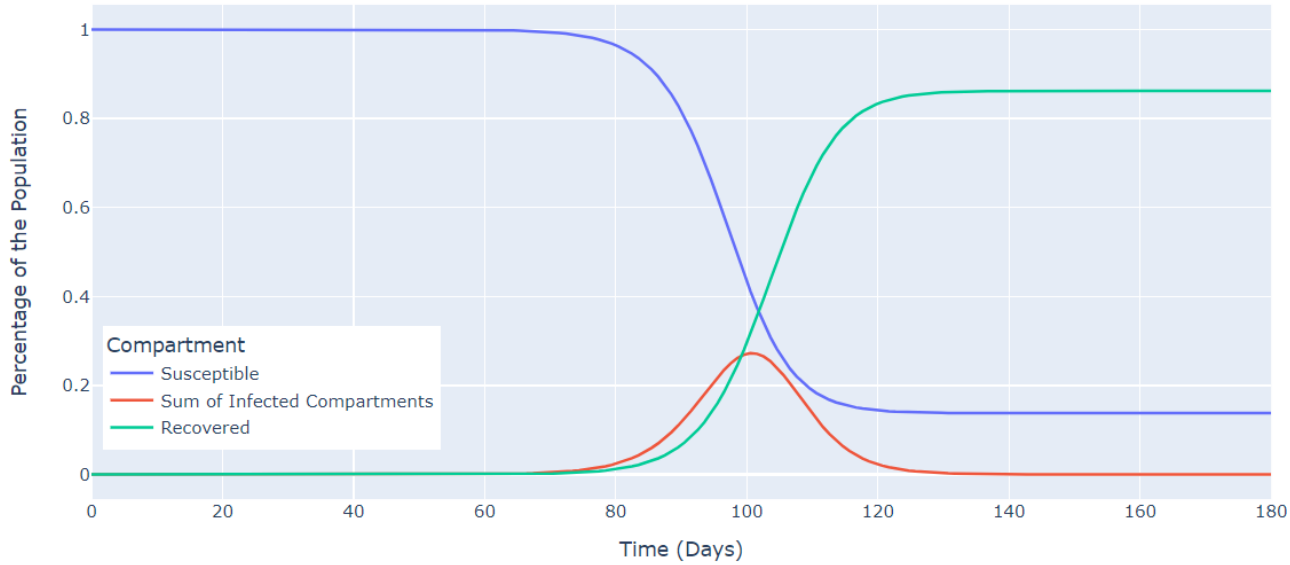

Figure 23: $SI_{10}R$ Model of Log-normal Distributed Generation Interval

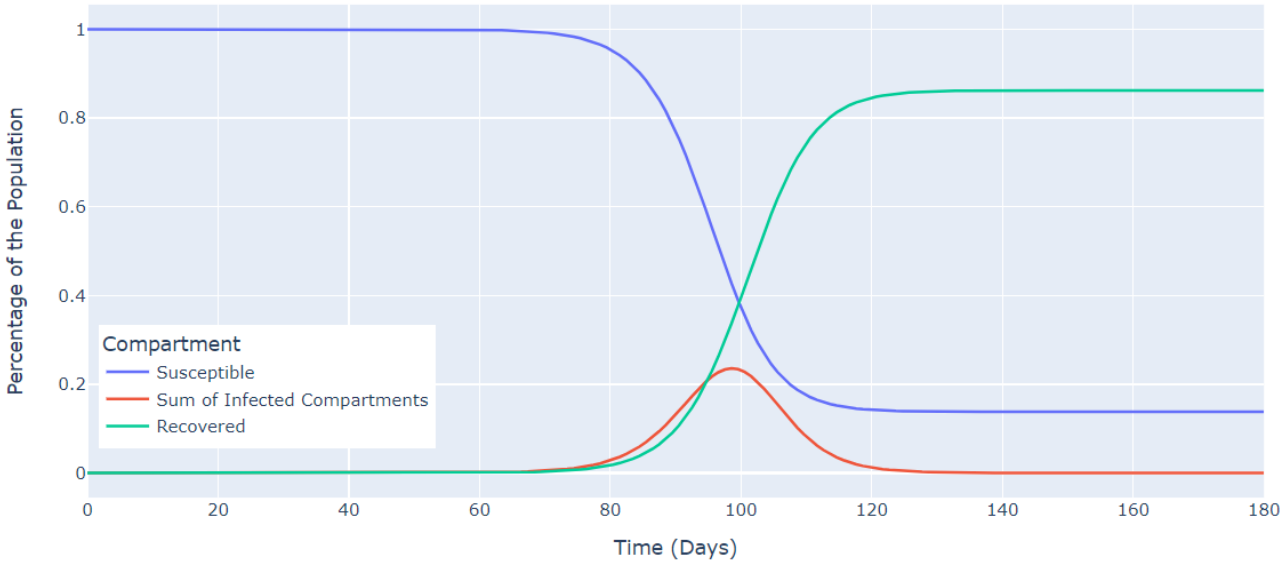

Figure 24: $SI_{10}R$ Model of Weibull Distributed Generation Interval

B $SI_{10}RS$ Compartmental Model Results for Gamma, Weibull, and Log-normal

Below are each of the results for the $SI_{10}R$ model in their own figure. To compare each of these directly, see Figure 20.

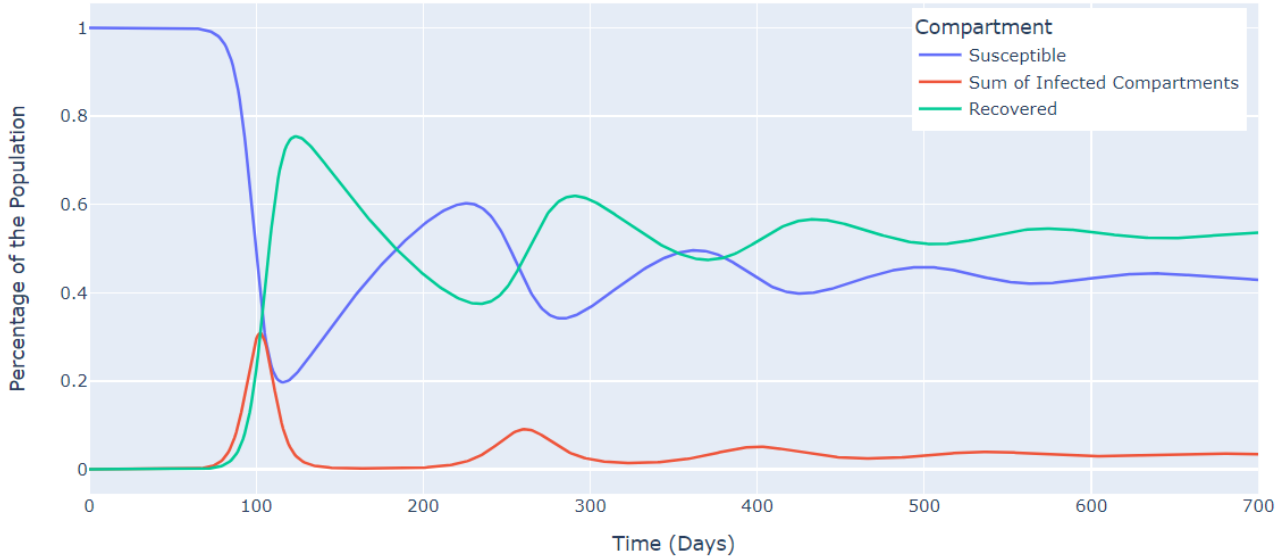

Figure 25: $SI_{10}RS$ Model of Gamma Distributed Generation Interval

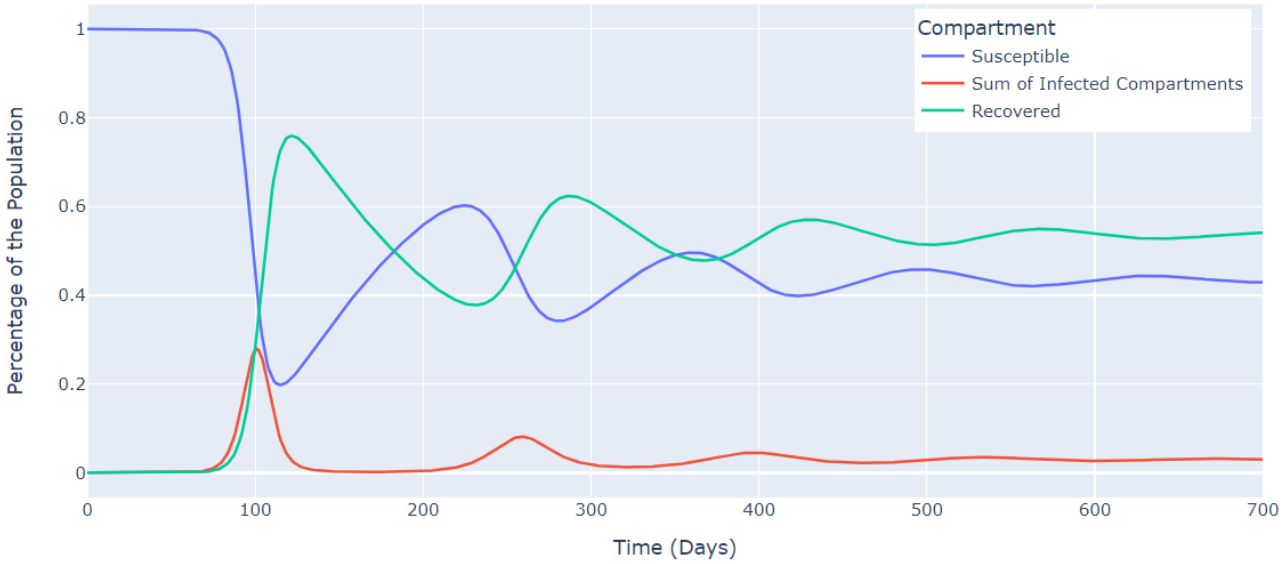

Figure 26: $SI_{10}RS$ Model of Log-normal Distributed Generation Interval

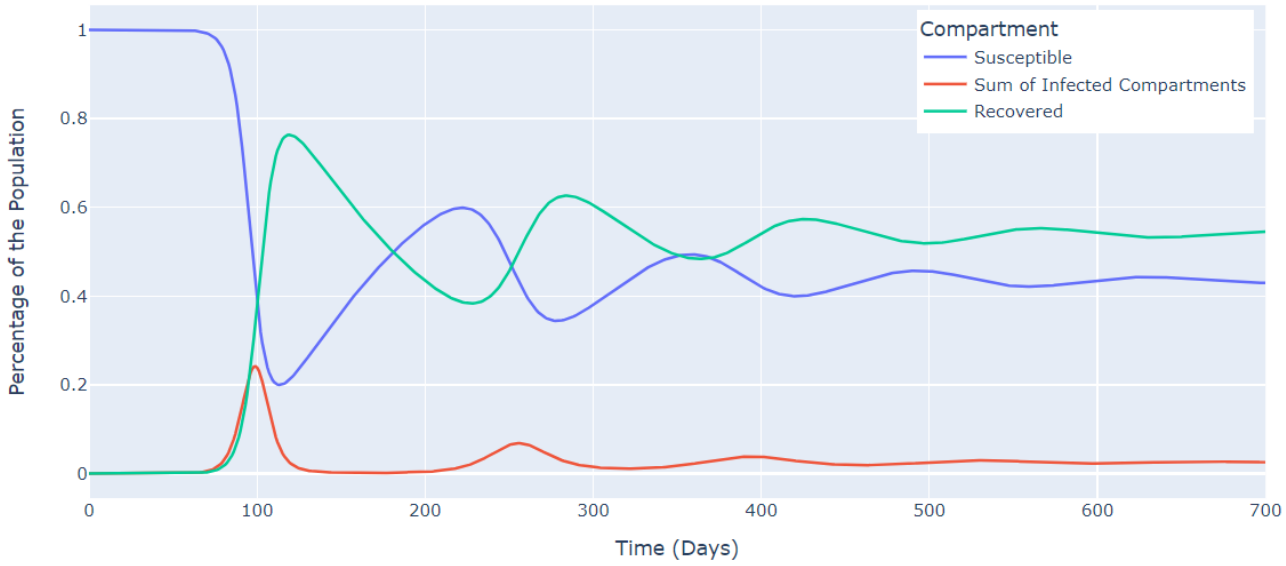

Figure 27: $SI_{10}RS$ Model of Weibull Distributed Generation Interval